Pleasing enhances Indirect Reciprocity based Cooperation under Private Assessment

Marcus Krellner[1,2] and The Anh Han[1]

[1] School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough, UK TS1 3BA

[2] Technische Universität Dresden, Germany

Emails: m.krellner@tees.ac.uk, t.han@tees.ac.uk

Abstract

Indirect reciprocity is an important mechanism for promoting cooperation among self-interested agents. Simplified, it means you help me, therefore somebody else will help you (in contrast to direct reciprocity: you help me, therefore I will help you). Indirect reciprocity can be achieved via reputation and norms. Strategies, such as the so-called leading eight, relying on these principles can maintain high levels of cooperation and remain stable against invasion, even in the presence of errors. However, this is only the case if the reputation of an agent is modeled as a shared public opinion. If agents have private opinions and hence can disagree if somebody is good or bad, even rare errors can cause cooperation to break apart. We show that most strategies can overcome the private assessment problem by applying pleasing. A pleasing agent acts in accordance with others' expectations of their behavior (i.e. pleasing them) instead of being guided by their own, private assessment. As such, a pleasing agent can achieve better reputations than previously considered strategies when there is disagreement in the population. Pleasing is effective even if only the opinions of few other individuals are considered and when it bears additional costs. Finally, through a more exhaustive analysis of the parameters' space than previous studies, we show that some of the leading eight still function under private assessment, i.e. that cooperation rates are well above an objective baseline. Yet, pleasing strategies supersede formerly described ones and enhance cooperation.

Keywords: Evolution of cooperation, reputation, indirect reciprocity, evolutionary game theory, private assessment.

1 Introduction

Cooperation is already considered central to complex life (Nowak, 2006; Perc et al., 2017) and may prove central to artificial systems as well (Langton, 1997). But cooperation relies on a handful of mechanisms that can enable it (Perc et al., 2017; Nowak, 2006). These include kin (Hamilton, 1964) and group selection (Traulsen and Nowak, 2006), direct and indirect reciprocity (Nowak, 2006; Nowak and Sigmund, 2005), spatial networks (Szabó and Fáth, 2007; Perc et al., 2013), different forms of incentives (Fehr and Gachter, 2000; Sigmund et al., 2001), and pre-commitments (Nesse, 2001; Han et al., 2013). Among these, indirect reciprocity stands out because it does not require any relatedness or other structural order between individuals, nor does it require repeated interactions (Rand and Nowak, 2013; Nowak and Sigmund, 2005; Okada, 2020a). This makes it especially relevant for modern global economies, where human or artificial agents often engage in non-repeated interactions. Indeed, real-world applications of indirect reciprocity, particularly, reputation-based systems, are prevalent in e-commerce (Jiang and Li, 2007; Standifird, 2001), socio-technical systems (Andras et al., 2018; Jones et al., 2013) and artificial societies (Conte and Paolucci, 2002; Smaldino and Lubell, 2014).

Indirect reciprocity is usually modeled using self-interested agents playing the donation game (Sigmund, 2016): a random agent (donor) is selected to pay a personal cost c to grant benefit b to another randomly selected agent (recipient). The benefit is assumed to be bigger than the cost, i.e. $b > c$. Hence, it is best for the population collectively if every agent decides to take the cost, so that the sum of all wealth would increase by $b - c > 0$ with each interaction. However, the individually preferred choice for each agent is to avoid the cost (i.e. defect), making the game a social dilemma. Thus, to maintain cooperation and prevent defection, agents may use strategies that are based on reputations and norms (Nowak and Sigmund, 1998a; Ohtsuki and Iwasa, 2006).

The reputation of the recipient determines how the potential donor should act (pay the cost or not). Norms determine what reputations the acting agent will earn. A simple version of such a strategy may state: If the reputation of the recipient is good, then donate, otherwise defect; and: if an agent donates, he earns a good reputation, otherwise a bad one (Nowak and Sigmund, 1998b). It was shown that such simple strategies, whose norms only consider actions, cannot maintain stable cooperation. With these norms, agents will earn a bad reputation if they do not donate to those with a bad reputation. This is known as the problem of justified punishment (Panchanathan and Boyd, 2003).

This problem can be solved using norms that also consider the current reputation of the recipient (so called second-order norms) or even the reputation of the donor itself (third-order norms) (Ohtsuki and Iwasa, 2004, 2006). Each possible scenario (a good/bad donor cooperates/defects against a good/bad recipient), linked to a resulting reputation (good/bad), leads to 256 possible norm combinations. An exhaustive search for successful norms showed that only eight can reliably maintain cooperation (Table 1). They are collectively referred to as the "leading-eight" but some are mentioned separately in the literature, e.g. L1 "standing" (Leimar and Hammerstein, 2001; Panchanathan and Boyd, 2003), L6 "stern judging" (Pacheco et al., 2006; Santos et al., 2018) and L7 "staying" (Okada et al., 2017a; Sasaki et al., 2017). Standing was one of the first strategies to address the problem of justified punishment (see above), whereas stern judging and staying have been suggested as the best indirect reciprocity norms in

<table>
<tr><th colspan="2">strategy</th><th colspan="8">assessment rules (norms)</th><th>situation</th><th colspan="4">action rules</th></tr>
<tr><td rowspan="3">name</td><td rowspan="3">number</td><td>G</td><td>B</td><td>G</td><td>B</td><td>G</td><td>B</td><td>G</td><td>B</td><td>donor</td><td>G</td><td>B</td><td>G</td><td>B</td></tr>
<tr><td colspan="4">C</td><td colspan="4">D</td><td>action</td><td colspan="4">(does not apply)</td></tr>
<tr><td colspan="2">G</td><td colspan="2">B</td><td colspan="2">G</td><td colspan="2">B</td><td>recipient</td><td colspan="2">G</td><td colspan="2">B</td></tr>
<tr><td>standing</td><td>L1</td><td rowspan="8">G</td><td rowspan="8">G</td><td>G</td><td>G</td><td rowspan="8">B</td><td rowspan="8">B</td><td rowspan="8">G</td><td>B</td><td rowspan="8"></td><td rowspan="8">C</td><td rowspan="8">C</td><td rowspan="8">D</td><td>C</td></tr>
<tr><td></td><td>L2</td><td>B</td><td>G</td><td>B</td><td>C</td></tr>
<tr><td></td><td>L3</td><td>G</td><td>G</td><td>G</td><td>D</td></tr>
<tr><td></td><td>L4</td><td>G</td><td>B</td><td>G</td><td>D</td></tr>
<tr><td></td><td>L5</td><td>B</td><td>G</td><td>G</td><td>D</td></tr>
<tr><td>stern judging</td><td>L6</td><td>B</td><td>B</td><td>G</td><td>D</td></tr>
<tr><td>staying</td><td>L7</td><td>G</td><td>B</td><td>B</td><td>D</td></tr>
<tr><td></td><td>L8</td><td>B</td><td>B</td><td>B</td><td>D</td></tr>
</table>

Table 1: Norms and action rules of the leading-eight. They show many similarities that often confirm with intuitive moral judgement, e.g. if a donor is currently considered bad and he cooperates with a good recipient, he can earn a good reputation, i.e. being forgiven. C = cooperates, D = defects, G = good reputation, B = bad reputation

recent years.

Most previous studies on the subject, however, assumed a simplified condition: public assessment (e.g. Leimar and Hammerstein (2001); Nowak and Sigmund (1998a); Ohtsuki and Iwasa (2004); Pacheco et al. (2006); Panchanathan and Boyd (2003); Santos et al. (2018); Suzuki and Kimura (2013); Xu et al. (2019)). It means, that every agent has a single reputation value (good or bad), which is agreed upon by all. This value may be wrong, i.e. a cooperative player may have a bad reputation because an error had occurred, but this error will be unanimously shared. One way to interpret such a public reputation system is a scenario where a single agent observes the interaction between a donor and a recipient, upon which he might commit an error, and then shares his perception with all others. If however, at least two agents observe the interaction and may commit perception errors independently, they may disagree on the donor's reputation afterwards. Thus, instead of a single public reputation, there are private opinions. This causes severe problems to the reputation based moral system since the current reputation (or opinion) influences the assessment of future interactions. Two agents with different opinions may assess the same situation differently, even if no further errors occur (see an example shown in Figure 1). This way, with enough initial perception errors, bad opinions can wrongly cascade through the population, even to the point where no one will cooperate any longer. Stern judging is especially struck (Okada et al., 2017a; Brandt and Sigmund, 2004) and some have even argued that the problem of private assessment (aka private monitoring or private information) is detrimental for all leading-eight (Hilbe et al., 2018). Yet, there are also studies which suggest that at least staying is doing okay, in simulations (Okada et al., 2017a) as well as via analytical approaches (Okada et al., 2018, 2020).

In this paper, we investigate an approach to coping with private assessment, first mentioned in (Krellner and Han, 2020), called 'pleasing'. Pleasing rests on two simple rationales: i) agents are being aware that others may have distinct opinions and ii) agents want to maximize the chance of getting a good reputation (so as to obtain a better payoff). For instance, if a donor wants to defect, but the majority of

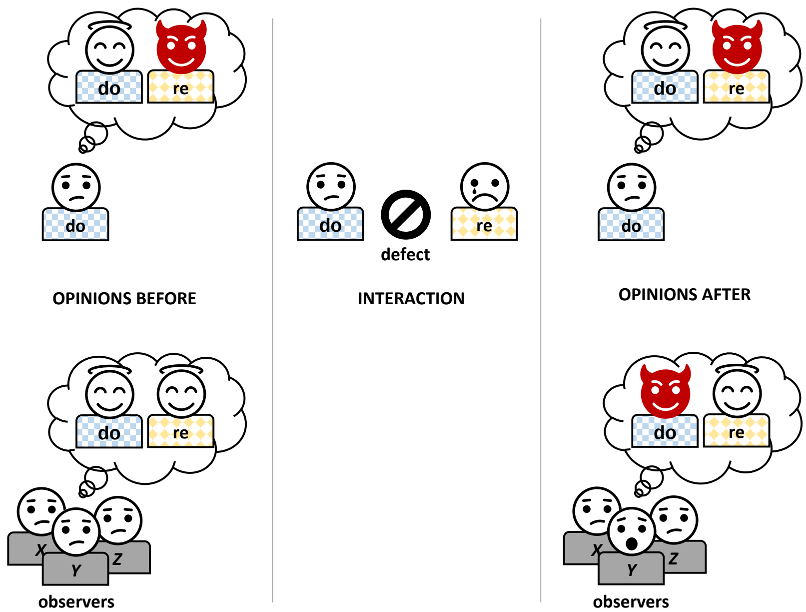

Figure 1: Example of a disagreement error. Left column shows opinions before the interaction. Only the opinion about the recipient (re) differs for the observers and the donor (do). Since the donor considers the recipient bad, he refuses to help. For the observers, who consider the recipient as good, this appears as a bad act. Hence the reputation of the donor suffers and disagreement spreads.

players think he should cooperate (Figure 1), he will cooperate to please them.

In this scenario, information sharing is not a collective act. Instead only the current donor is actively seeking information. Which seems almost natural for him to do, since this information is viable for his future reputation. This enables one to avoid a major problem associated with information sharing: who is responsible and ready to pay the cost of the information distribution (Suzuki and Kimura, 2013)? It is natural to assume that the direct beneficiary of some information would be willing to make effort to gather it. The donor could even decide to compensate others for the information they share, depending on how much additional cost pleasing is worth to him.

Is it possible to solve the problems of private assessment with 'pleasing'? Judging from moral intuition, pleasing may seem an undesirable behavior. People who are too easily persuaded by the majority might be looked down upon. And it is easy to imagine situations where pleasing would cause harm to society or individuals; when one goes along with punishing people who do not deserve it. But are such acts the misfire of an otherwise good mechanism that prevents us more often from punishing people we ourselves mistakenly believe to deserve it? And besides human societies, can pleasing inform how artificial agents can be designed and engineered to maximize cooperative behaviour and overall productivity (Paiva et al., 2018; Andras et al., 2018)? We contribute to answering these questions and to the question, whether indirect reciprocity does work under privates assessments. We therefore study pleasing as well as original leading eight, from a strict evolutionary game theory perspective (Sigmund, 2016) in wide parameter ranges.

The remainder of the paper is structured as follows. We next describe our models and methods. We then show how pleasing improves the evolution of cooperation first under the assumption that mutations are rare (Hilbe et al., 2018; Sigmund, 2016; Han et al., 2013), to compare and build upon the results of Hilbe et al. (2018) and Krellner and Han (2020), where the same assumption was adopted. We then analyze how robust pleasing is if only a few other agents (instead of all possible agents) are pleased and when pleasing bears an additional cost. Thirdly, we examine the influence of selection strength under rare mutations. Then, relaxing the assumption of small mutations, we study the impact of varying both selection strengths and mutation rates. Finally, we discuss the implications of our findings while pointing out limitations and future directions. We also provide an appendix with additional results, to complement those in the main text and show the robustness for different parameter choices.

2 Models and Methods

We test pleasing for its evolutionary stability and cooperativeness. We run agent-based simulations under private, noisy and incomplete information, adopting a similar setting as in Hilbe et al. (2018); Krellner and Han (2020), for a clear and convenient comparison. We consider four strategies in co-presence: two unconditional strategies, i.e. always cooperate (AllC) or always defect (AllD); and two conditional strategy based on norms, i.e. the original leading eight (O_1, O_2, ... $\in \mathrm{O}_x$) and the pleasing leading eight (P_1, P_2, ... $\in \mathrm{P}_x$). We also compare the performance of pleasing and the original leading eight when playing against AllC and AllD, i.e. when there are three strategies in co-presence, in order to provide a direct comparison with the analysis in Hilbe et al. (2018) (section 3.4). The original leading eight behave as described in the literature (Table 1). P_x players may also try to please other P_x agents.

We use two complementary methods to study stability and resulting levels of cooperation. Both build upon the results of Krellner and Han (2020), who studied three strategies at a time under the limit of rare mutation. First, we put all four strategies together under the same restriction. This way we can provide a direct competition between the previously studied leading-eight strategies, O_x, and the newly defined pleasing ones, P_x, and how their struggle influences the evolution of cooperation as a whole. Second, we relax the restricting assumption to study evolutionary dynamics under any mutation rate. For this second approach, we consider again only three strategies in co-presence: AllC, AllD and either O_x or P_x. This way we can readily compare our results to other previous simulations (Okada et al., 2017b; Hilbe et al., 2018) and to previous theoretical analyses (Okada et al., 2018; Okada, 2020b). Using three strategies keeps interpretations clear and concise and reduces the burden of computational demand (which can be very excessive) [1]. In both experiments, we analyze stability as the frequency of the strategy under evolutionary pressure. We subsequently compute the average cooperation rate in the evolving population (denoted by **Copop**), which we use to indicate the success of indirect reciprocity.

[1]The computation demand increases exponentially with each additional strategy. For example with $N = 50$, a single parameter setting of two strategies requires 51 simulations, three strategies 1275 and four strategies would require 22100.

2.1 Model and Agent-based Simulation Setup

We consider a well-mixed population of size N. Reputations have two possible states, 1 (good) or 0 (bad), and players can choose between two actions, cooperate or defect. Agents have either an unconditional strategy (AllC, i.e. always cooperate, or AllD, i.e. always defect) or a conditional one. Conditional strategies are defined by a vector pair (α, β) that represent their assessment and action rules. The assessment rule α has eight entries, one for each combination of three situational features: action and current reputations of donor and recipient. An entry is equal to 1 if the resulting reputation is good, otherwise it is 0. The action rule β corresponds to the four possible combinations of reputations of the donor and recipient. An entry is 1 if the agent will cooperate in that situation and 0 if he will defect. As in Hilbe et al. (2018); Krellner and Han (2020), we will focus on the 'leading-eight' strategies L_x (Ohtsuki and Iwasa, 2006) (see Table 1).

2.2 Reputation Dynamics

We simulate the reputation dynamics for players in a population with at most three fixed strategies. The state of reputation is given by the $N \times N$ image matrix $M(t)$ of the population at time t (Uchida, 2010). An entry $m_{i,j}(t)$ is equal to 1 if player i has a good opinion of player j, and 0 otherwise. Initially all entries are set to 1, a state of homogeneously good reputation. To ensure that initial image matrix has no effect, only data after an initial transition period is used to compute results. As described in Hilbe et al. (2018), this way other initial states only change results for the extreme opposite, homogeneously bad reputations, and even then, only for L_7.

Each time step in a simulation consists of three parts. First, a donor do and a recipient re are drawn at random from the population. The donor then decides whether to cooperate. Unconditional players always act the same, whereas the standard leading-eight players decide by their action rule β and their opinions of themselves $m_{do,do}(t)$ and of the recipient $m_{do,re}(t)$. For pleasing leading-eight players see below. In the third step, reputations are updated due to observations.

The donor and the recipient always observe the interaction, whereas other players independently observe it with probability q. Any observation (even by the focal donor or recipient) is independently altered due to error with probability ϵ towards the opposite action (e.g. cooperation instead of defection). Each player who has observed the interaction updates her opinion of the donor according to her assessment rules, the action she observed and her current opinions (i.e. even the donor updates her opinion of herself). The result is the image matrix $M(t+1)$. This process is repeated for 10^6 rounds in all the results shown below).

2.3 Pleasing

Pleasing P_x discriminators differ from the original O_x discriminators in two ways. They share their opinion with others (O_x won't). Also, they decide how to act based on the opinion of the majority when being chosen as a donor. Let n_w be the number of players the donor wants to please and N_P the number of pleasing P_x players in the current population. The donor will consider the opinions λ of n_p pleasing players, where $do \notin \lambda$ and n_p is the minimum of n_w and N_P. The average opinions $\widehat{m}_{do}$ and $\widehat{m}_{re}$ are given by

$$\widehat{m}_y = \frac{m_{do,y} + \sum\limits_{i=\lambda} m_{i,y}}{n_p + 1}. \tag{1}$$

The majority opinion

$$m_{maj,y} = \begin{cases} 0, & \text{if } \widehat{m}_y < 0.5, \\ 1, & \text{if } \widehat{m}_y > 0.5, \\ \text{randomly 0 or 1,} & \text{otherwise.} \end{cases} \tag{2}$$

The pleasing donor will act according to her action rule as if the majority opinion was her own. Afterwards observers assess the action as before (Note that the averaged opinions are not used again, especially not to update the opinions of the donor. AllC and AllD do not have opinions, O_x do have, but neither are sharing information.)

2.4 Payoffs

Based on simulations we compute the average rate of cooperation $x_{A,B}$ of agents with strategy A towards agents with strategy B. To ensure that initial image matrix has no effect on these values, we disregard the first half the simulation, so out of 10^6 total time steps we only use the data of the second half to compute cooperation rates, similar to (Hilbe et al., 2018). There are q strategies present in a population of size N, each is played by n_i agents in a population of $N = \sum_{i=1}^{q} n_i$. The average payoff per interaction of an agent with strategy A (denoted by $\widehat{\pi}_A$), in which the agent is involved, is given by the benefits she receives minus the cost. It can be written as follows

$$\widehat{\pi}_A = \frac{b}{N-1} \sum_{i=1}^{q} x_{i,A} n_i^* - \frac{c}{N-1} \sum_{i=1}^{q} x_{A,i} n_i^*, \qquad \text{where } n_i^* = \begin{cases} n_i - 1, & \text{if } A = i, \\ n_i, & \text{otherwise.} \end{cases} \tag{3}$$

For some analysis we also subtract a cost c_p for pleasing strategies. Thus, the payoff for a leading-eight strategy that adopts pleasing A_p is given by $\widehat{\pi}'_{A_p} = \widehat{\pi}_{A_p} - c_p$.

2.5 Social learning through imitation

We assume that players may consider updating their strategy according to their accumulated payoffs. We model this process using Evolutionary Game Theory methods for finite populations (Traulsen et al., 2006). Particularly, in each step of the evolutionary process, two players, i and j, are randomly selected from the population for update via pairwise comparison: i will adopt the strategy of j with a probability given by the Fermi function

$$p_{i,j} = (1 + e^{-s(\widehat{\pi}_j - \widehat{\pi}_i)})^{-1}. \tag{4}$$

The parameter s defines the 'imitation strength' or 'selection strength', i.e. how much p depends on the payoff difference. For $s = 0$, imitation will be entirely random and as s approaches infinity the imitation process becomes more deterministic. In most of the results presented below, we adopt $s = 1$ for a fair and convenient comparison to Hilbe et al. (2018), but we also study the influence of different regimes of selection strengths.

2.6 Evolutionary Dynamics under the Limit of Rare Mutation

We want to test the stability of pleasing strategies P_x against all possible rivals. Thus, for each pleasing strategy we consider populations with players that either play the P_x strategy at hand, its original version O_x, AllC or AllD, so four strategies in total. We further assume for simplicity and convenience the limit of rare mutations (Nowak et al., 2004; Wu et al., 2012). A mutation is an event where a player of strategy A adopts any of the remaining strategies at random. In the limit of rare mutation, the population will almost always be in a homogeneous state with a single strategy present, since these states are absorbing and cannot be escaped by imitation alone. If a population consists of only strategy B and a new mutant of strategy A arrives, the population will subsequently either reach a homogeneous state of A or B and there will be at most these two strategies present at during the transition. The probability that A will take over, i.e. fixate, is denoted by $\rho_{B,A}$. It is determined through the probabilities of incremental steps towards more players of A, T_A^+, or the reverse step T_A^-, as follows.

Let k $(1 \leq k \leq N-1)$ be the number of individuals using strategy A and $(N-k)$ the number for strategy B, in the population, then

$$T^+(k) = \frac{N-k}{N}\frac{k}{N}p_{A,B}. \tag{5}$$

Then,

$$\rho_{B,A} = \big(1 + \sum_{i=1}^{N-1}\prod_{j=1}^{i}\frac{T^-(j)}{T^+(j)}\big)^{-1}. \tag{6}$$

These fixation probabilities determine a Markov Chain that describes the evolutionary transitions between the homogeneous states of q strategies, with a transition matrix Z of size $q \times q$. The entry (of matrix Z) $z_{A,B} = \rho_{B,A}/(q-1)$ for $A \neq B$ and $z_{A,A} = 1 - \sum_{j=1}^{q} z_{j,A}$. The normalized eigenvector with eigenvalue 1 of the transposed matrix of Z provides the selection-mutation equilibrium σ for each considered strategy. It describes, in the long run, how often the population will spend in a homogeneous state of the corresponding strategy (Fudenberg and Imhof, 2006). (In some cases a unique eigenvector cannot be computed numerically. In such cases we would simulate the times spent in homogeneous states with the transition matrix as follows. Start in each strategy state, randomly transition according to matrix for 10^6 times while add up instances of being in a state, add all q runs to gain σ.) This small mutation limit approach has been shown to be widely applicable to scenarios which go well beyond the strict limit of very small mutation rates (Sigmund et al., 2010; Rand et al., 2013; Zisis et al., 2015).

Evolutionary success of discriminators such as P_x or O_x is not a guarantor for high cooperation rates. Strategies may succeed with little or no cooperation as players adopting such strategies defect against each other in their homogeneous state (e.g. P_8 and O_8). We therefore simulate the cooperation rates within each homogeneous population. We then obtain the average cooperation $\widehat{X}$ (Copop) of the whole system $\widehat{X} = \sum_{i=1}^{q} \sigma_i x_{i,i}$, where σ_i is the frequency of strategy i at the selection-mutation equilibrium and $x_{i,i}$ the cooperation rate in the homogeneous state of strategy i (which is one for AllC and zero for AllD).

2.7 Evolutionary Dynamics under any Mutation Rate

Despite the usefulness of the approximation in the limit of rare mutations, it presents unavoidable limitations to the dynamics of a population as described above, e.g. that only two strategies can be present at the same time. Moreover, it has been shown experimentally that behavioral mutations or exploration occur frequently in human interactions (Traulsen et al., 2010; Rand et al., 2013). Non-rare mutations also play an important role in enabling cooperation in the context of social dilemmas (García and Traulsen, 2012; Duong and Han, 2020; Antal et al., 2009). Hence we consider evolutionary dynamics under any mutation rate between $q = 3$ strategies: AllC, AllD and a discriminator, either O_x or P_x. In such a setting, the finite population of size N can reach any composition state given by the set

$$\Delta_N^q := \{n = (n_1, n_2, n_3)|\ 0 \leq n_i \leq N, \sum_{i=1}^{q} n_i = N\}. \tag{7}$$

There are $|\Delta_N^3| = \frac{(N+2)(N+1)}{2}$ possible population composition states. A transition between two such states is only possible if they are neighbors in that graph, i.e. if the number of agents of strategy A decreases by 1 and the number of agents of strategy B increase by 1 and both are possible. Such a transition can occur as a result of a mutation or imitation. A mutation occurs with probability μ. Hence the transition probability between state X and Y, which constitutes an decrease in A and increase in B, is

$$w_{X,Y} = \frac{n_A}{N}\left(\frac{\mu}{q-1} + (1-\mu)\frac{n_B}{N-1} p_{A,B}\right) \tag{8}$$

i.e. the probability, to pick an agent of strategy A, times the added probabilities of a mutation towards B and an imitation of B. Imitation only happens if there was no strategy change due to mutation. The probability of this specific imitation depends on the frequency of B and the payoff differences between the two (see equations 3 & 4).

With $q = 3$, a state has at most six possible neighbors to transition to. The probability to transition to any other state is 0. The probability to stay in the current state is 1 minus the sum of all transition probabilities. All these probabilities together define a Markov Chain of all population states with a transition matrix of size $|\Delta_N^3|$ x $|\Delta_N^3|$. For small enough population size, N, and number of strategies q, we can again compute the normalized eigenvector with eigenvalue 1 of the transposed matrix and obtain the selection-mutation equilibrium σ.

Each of the $|\Delta_N^3|$ components of the vector σ describes the time spent in a particular population state. We can therefore compute the relative frequency of the strategies by multiplying the number of agents in the particular state, as follows

$$f_A = \frac{1}{N} \sum_{i=1}^{|\Delta_N^3|} \sigma_i n_{i,A}. \tag{9}$$

Note that this formula can also be applied to the rare mutation limit that was described above, where there are only q homogeneous states, $n_i = N$ and $f_A = \sigma_A$.

Furthermore, we can compute the average cooperation rate $\widehat{X}$ (Copop) with the average cooperation in each state X

$$\widehat{X} = \sum_{i=1}^{|\Delta_N^3|} \sigma_i X_i. \tag{10}$$

The average cooperation in a state is computed with the average cooperation of strategy j towards k and their quantities in that state

$$X_i = \frac{1}{N(N-1)} \sum_{j=1}^{q} n_j \sum_{k=1}^{q} n_k^* x_{j,k} \quad \text{where } n_k^* = \begin{cases} n_k - 1, & \text{if } j = k \\ n_k, & \text{otherwise} \end{cases}. \tag{11}$$

3 Results

3.1 Improvements by Pleasing

We study the evolutionary dynamics between four strategies: P_x, O_x, AllC and AllD. The key parameters to be considered are the error rate ϵ (in this study an observation error) and the benefit size b (the cost c is normalized to 1). The benefit determines whether cooperation is worth the effort in the first place. Only $b > 1$ allows cooperators to outperform defectors. However, AllC will always be invaded under rare mutations by AllD for any b, since the former always pays the cost in addition to receiving the same benefits as the latter. Only discriminators such as P_x and O_x can actually outperform defectors by withholding benefits to them but sharing them amongst themselves. This act requires an additional capability for information acquiring and processing (even in a simple form), which is why the noise in acquiring of correct information, i.e. the error rate, is the second important factor.

We first investigate P_x when pleasing is not restricted (i.e. players can acquire and consider the opinion of all other pleasing players) for varying ϵ and b (fixing $c = 1$). Results for Copop are shown in Figure 2. The abundance of warm colors in the right columns (i.e. with pleasing) compared to the left ones shows the great potential of the new pleasing strategies. Copop is increased in wide parameter spaces for most L_X. With pleasing, higher benefits almost always increase cooperation rates and high Copop is even possible for frequent errors. P_1, P_6 and P_7 utilize pleasing the best among the leading-eight. High benefits and low error rates (lower right quadrant) seem generally best for cooperation, compared to low benefits and high error rates (upper left quadrant).

The improvements are the result of both pleasing strategies' high abundance and high cooperation in homogeneous state (Coho). Coho of most P_x are in fact close to 100% for $\epsilon < 0.3$, see Figure 3. The

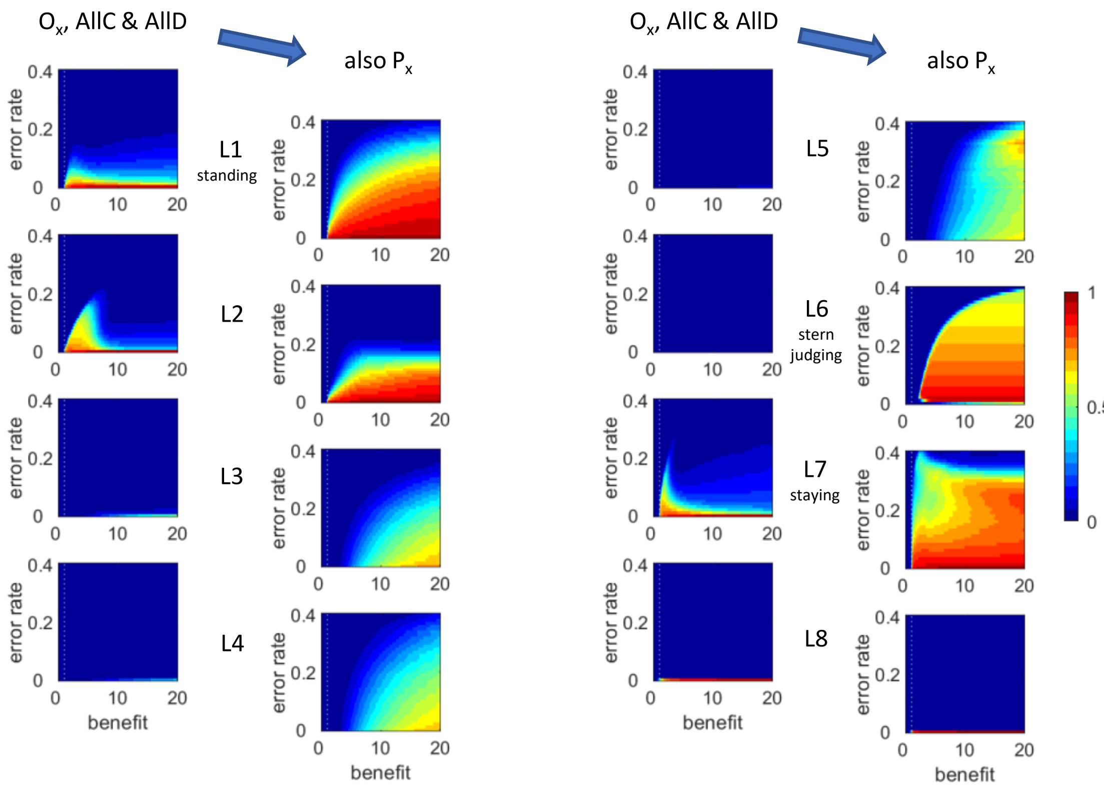

Figure 2: Effect of pleasing on the evolution of cooperation (Copop), across benefit b (x-axis) and error rate ϵ (y-axis). First and third columns show Copop in populations of original leading-eight O_x, AllC and ALLD; second and fourth columns show populations augmented by a fourth strategy P_x. Parameters: $N = 50$, $n_w = 50$ (i.e. pleasing all P_x players in the population), $q = 0.9$, $s = 1$.

General effect: Pleasing greatly facilitates the evolution of cooperation in large parameter regions. Copop is lowered by pleasing only in a small area of low error rates and low benefits for a single strategy family, L_2. For these parameters, pleasing slightly decreases the abundance of discriminators (both P_2 and O_2) and hence the resulting Copop.

Parameter space: The original leading-eight were successful to some extent only if error rates and benefits were both relatively low (lower left quadrant). More frequent errors but also higher benefits decrease their abundance (due to invasion by AllC). With pleasing, higher benefits almost always increase cooperation rates instead. Now, high benefits can also offset the negative effects of high error rates (lower right quadrant). Even if error occur as often as every third time, there are still a few cases that can maintain more than 50% cooperation.

Different strategies: With original leading-eight, only three of the leading-eight were successful to some extent (L_1 "standing", L_2 & L_7 "staying"), showing small parameter ranges that allow for at least 50% cooperation. With pleasing, this is true for all but one strategy, L_8. Although relatively abundant, L_8 showed no cooperation, with or without pleasing. Four populations (L_1, L_6 “stern judging” & L_7, less so L_2) can maintain more than 70% cooperation for large parameter ranges.

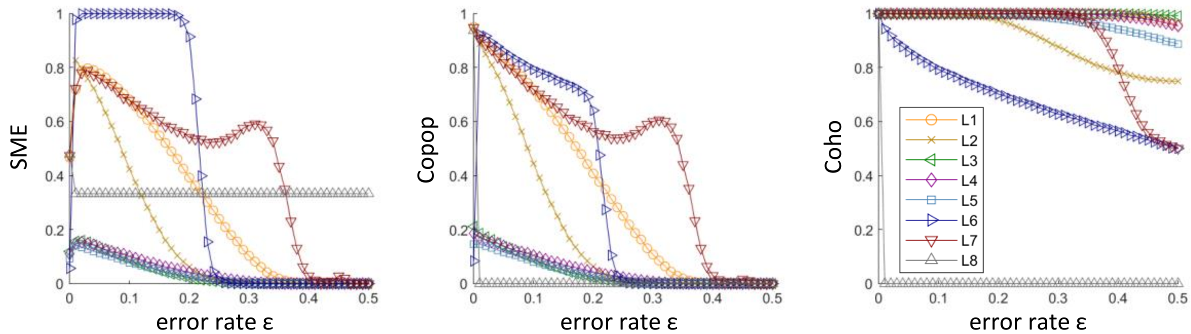

Figure 3: The abundance of P_x (i.e. SME, left side), their cooperativeness in homogeneous state (Coho_P, right side) and the resulting Evolution of cooperation in the whole population (Copop, center) for different error rates ϵ in populations of O_x, P_x, AllC and AllD. Other parameters: $N = 50$, $q = 0.9$, $s = 1$, $b = 5$. Since O_x and AllC are rare in almost all populations (Figure A1), Copop is roughly SME_P times Coho_P. Both are reduced by more frequent errors. However, for most strategies, such as P_1 and P_7, Coho remains maximal until almost every third observation is wrong ($\epsilon = 0.3$). In those cases, the bottleneck for Copop is the stability of the pleasing strategy. For P_6 on the other hand, stability remains maximal for some time whilst Coho steadily decreases. The bottleneck is the cooperativeness of the pleasing strategy (the same is true for P_8, which shows no cooperation at all). The rather unnatural condition of $\epsilon = 0$ shows some special dynamics. The cooperation in a P_6 population is greatly reduced, whereas it is the only condition in which the cooperation in a P_8 population is considerable.

differences in the resulting cooperation rates (Copop) were mainly caused by differences in frequency. The only exception, as in previous studies (Hilbe et al., 2018; Krellner and Han, 2020), was P_8, which failed to have any cooperation with itself and was therefore exactly as frequent as AllD. Pleasing Stern judging (P_6) is also different in that it still dominates the population almost all the time for increasing error rates but becomes steadily less cooperative.

The observation rate, the probability that a non-involved player perceives what the donor does, was previously reported to be of little effect when any errors are present (Brandt and Sigmund (2004), compare also with Hilbe et al. (2018)). The same is true for pleasing, except if zero observation rate is approached.

The current results of stability and cooperativeness are a replication of (Krellner and Han, 2020), in which a population consisted of only three strategies at a time. The inclusion of the original leading eight strategies in the populations of this study did not change the success of pleasing. It is however important to analyse this larger set of strategies together in co-presence. Adding (or leaving out) a strategy can strongly affect the final outcome of population dynamics. Strategies can catalyze each other, so a formerly weak one might be catalyzed by other strategies that were ignored in the first place, and become dominant, as observed in García and Traulsen (2019); Sigmund et al. (2010); Han (2016).

3.2 Evolution of Pleasing

To better understand this remarkable capability of pleasing to promote cooperation under private assessment, we examine the detailed evolutionary dynamics in the case of P_1 (for illustration see Figure 4).

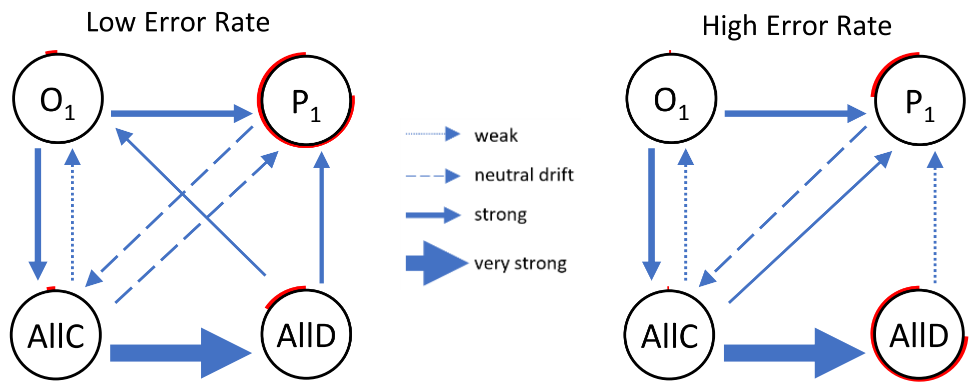

Figure 4: Fixation probabilities and frequency for populations of O_1, P_1, AllC and AllD for two conditions, left: $\epsilon = 0.05$, right: $\epsilon = 0.25$. In both conditions: $N = 50$, $q = 0.9$, $b = 5$, $s = 1$. Frequency is depicted as red surface around the circles representing the homogeneous population states. Fixation probability is depicted as follows: $0 \leq \rho < 0.001$ arrow omitted, $0.001 \leq \rho < 0.015$ weak, $0.015 \leq \rho < 0.025$ neutral drift (in the sense that $\frac{1}{N} = 0.02$ is the transition with exactly similar payoffs), $0.025 \leq \rho < 0.15$ strong and $0.15 \leq \rho$ very strong, transitions. Exact values are given in Table A2. The most important difference is the less likely transition from AllD to P_1 ($\rho = 0.07$ for rare errors, $\rho = 0.005$ for frequent errors).

Pleasing is risk dominant against all other strategies in both conditions (i.e. the transition probabilities towards it are always higher than those away from it, see also Figure A3). However, it is only the most frequent strategy for low error rates, whereas AllD is so for high ones. Even though, AllD is losing in direct comparison with P_1 in this condition.

To explain this unusual result, we look at a cyclic pattern between P_1, AllC and AllD under high error rates. Transitions from P_1 to O_1 or AllD are close to zero, so we ignore them for now. P_1 will transition to AllC with a chance of 0.02. That means, 298 out of 300 mutations fall back to P_1, the others transition to AllC. In there, a P_1 mutant has about 0.04 chance to cause a transition to P_1 (so higher than in return). An AllD mutant on the other hand has a 0.67 chance to cause a transition. I.e. in 4 out of 300 cases, AllC transitions back to P_1 (and will likely stay there for some time, increasing the frequency of P_1), but in 67 out of 300 cases it transitions to AllD. In AllD finally, only 0.5 of 300 mutations will lead to a transition towards P_1. Any other case means to stay in AllD, highly increasing its frequency (see also Table A2 and Figure A4).

In other words, although P_1 is risk dominant, the transition from P_1 to AllC and then to AllD is actually more likely than the transition from AllD back to P_1. O_1 plays no significant role, more so for high error rates, but practically in both conditions. The examples given is quite representative. We provide plots risk dominance across ϵ and b in Figure A3. A noticeable exception is P_6. It has a very low probability to invade AllD even for low error rates, but in turn no other strategies can invade it (i.e. probabilities approach zero, see Figure A5)

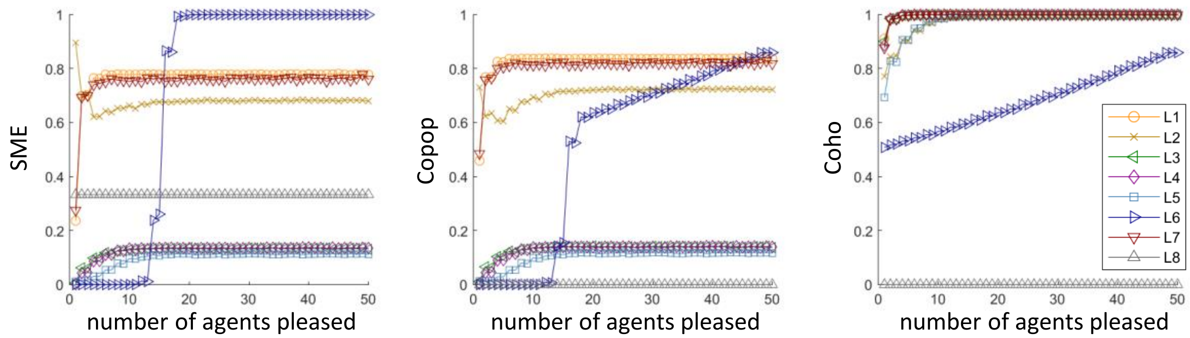

Figure 5: Number of agents pleased. Shown is the abundance of P_x (i.e. SME_P, left side), their cooperativeness in the homogeneous state ($Coho_P$, right side) and the resulting evolution of cooperation in the whole population (Copop, center), for different effort levels of pleasing, i.e. the number of players that the donor tries to please. Other parameters: $N = 50$, $q = 0.9$, $b = 5$, $\epsilon = 0.05$. As more players are considered, Copop often increases steeply and then levels of.

3.3 Efficiency of Pleasing

Enquiring all other opinions is likely to be expensive (e.g. calories, time or opportunity costs). We therefore analyze cooperation rates if players would try to please smaller numbers of agents. Figure 5 shows an example for some commonly used parameters, but our additional analysis show that the qualitative results are robust for other values of ϵ and b (see Figure A6). As the number of pleased players increases, Copop often increases steeply, then levels of. That means that pleasing a subset can be as effective as pleasing the entire population. Even pleasing just a few agents is enough for P_1 (standing) and P_7 (staying) to reach their full potential. P_6 (stern judging) on the other hand appears to have a threshold after which it rapidly reaches total stability (SME= 1), but each additional agent that it pleases makes it more cooperative. If error rates are higher, more agents are needed to be pleased, i.e. increases in Copop appear later or are less steep (Figure A6).

We also consider the cost of pleasing itself. Namely, pleasing players have to pay a cost c_p each time they are selected as donors. This cost represents the time and energy they spend obtaining others' opinions, but could also include compensation for some third-party agents from whom they gather information. Increasing the cost steadily decreases Copop with a single exception; P_6 (stern judging) was not affected by smaller costs (Figure 6, right side). Since costs do not change the cooperativeness in homogeneous states (Coho), the changes in Copop are entirely a result of lowered stability against the other strategies, which are not affected by costs. As described above, L6 had the highest SME and was stable enough to not lose any ground, even when baring considerable costs for pleasing. Additional analysis with different parameters support this finding, whereby higher benefits enable L6 to endure even higher costs (see Figure A7).

A noteworthy dynamics arises in the case of P_2. The Copop of its population first decreases with rising costs, before it increases again. Most populations follow the same pattern. With rising cost, pleasing players are steadily replaced by original discriminators and AllD, until they reach the same frequency as

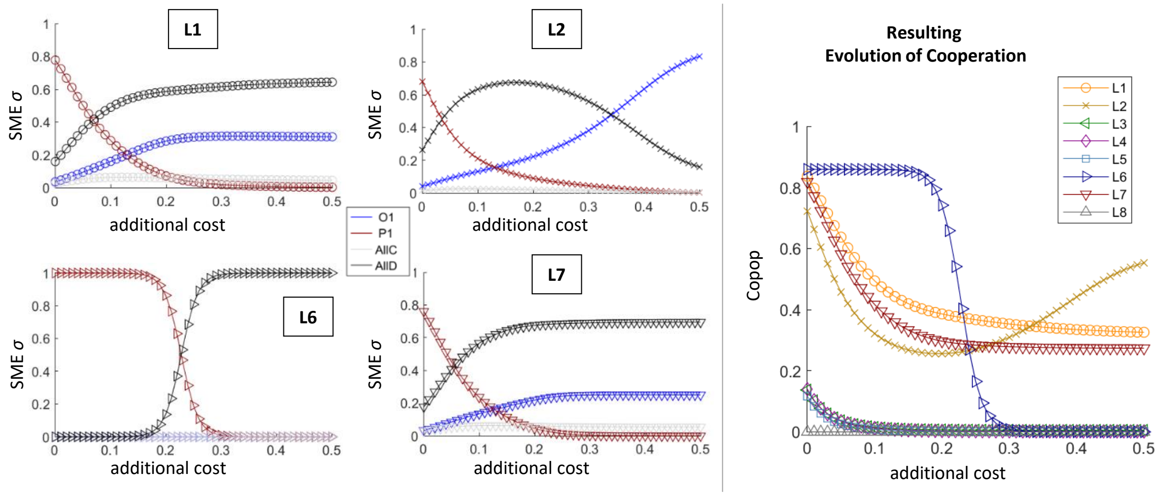

Figure 6: Dynamics across different additional costs of pleasing. Left: abundance (SME) of the four strategies O_x, P_x, AllC and AllD. Right: Cooperation in the population (Copop). Other parameters: $N = 50$, $q = 0.9$, $s = 1$, $b = 5$, $\epsilon = 0.05$. For the most part, increasing costs reduce Copop (right), until it reaches the level before the introduction of pleasing. The SME (left) shows, that pleasing strategies are gradually replaced by AllD and O_x, until the frequencies replicate the states for three strategies without P_x. In the case of P_6 this means, only AllD are left in the population and cooperation seizes. The exception to this pattern is P_2. Its frequencies is reduced, but the losses are first picked up by AllD. Only if costs rise even further, O_2 can reclaim frequency back from AllD and Copop increases again. P_6 (stern judging) is even entirely unaffected by small costs.

before the introduction of pleasing (Figure 6, left side). In the case of P_2 however, there seems to be an intermediate state, where pleasing agents are decimated, but original cannot yet take over. It lead to an intermediate rise of AllD in that population.

The results obtained here remain robust if we consider a population of three instead of four strategies (i.e. when removing the original leading eight) (Krellner and Han, 2020). Pleasing can be efficient and it can endure some additional cost, before losing its advantage.

3.4 Stochastic Effects: Varying Selection Strength and Mutation Rate

We now study the impact of varying the strength of selection s, i.e. how the evolutionary dynamics change on the continuum between random and almost deterministic. When considering selection strength, it is important to keep in mind, that very low s will allow any naive cooperator to persist in a population. If s is low enough, transitions do not depend on payoff differences at all but are just random drift. Hence, all strategies have the same frequency of $1/q$. So, at least under weak selection, the cooperation in the population depends heavily on the set of strategies that are present in the population of interest.

It is therefore necessary to discard low s as irrelevant or, as we suggest, to compare the dynamics

Comparison to Baseline across Selection Strength

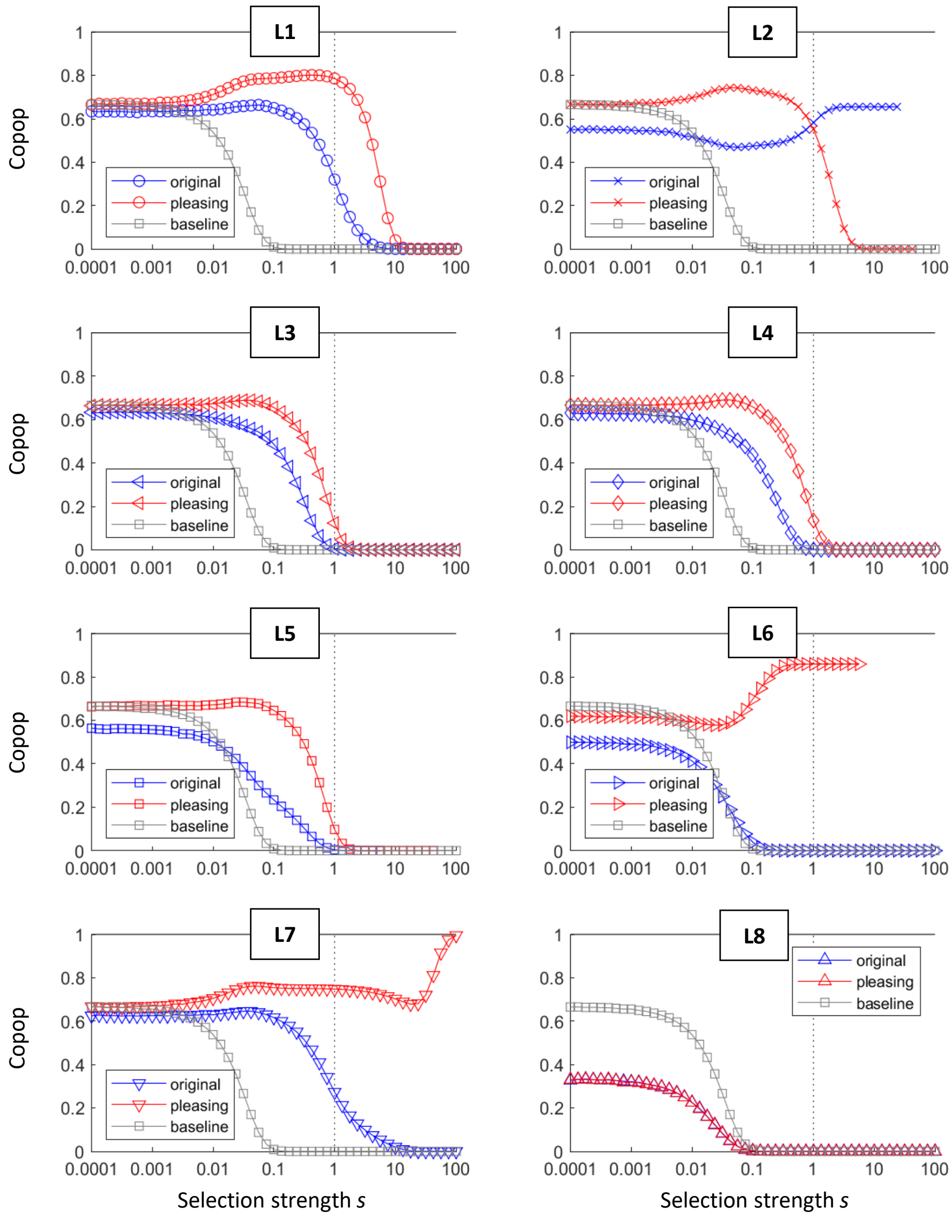

Figure 7: Evolution of Cooperation (Copop) across selection strength s in three different populations: original:={AllC,AllD,O_x}, pleasing:={AllC,AllD,P_x} and baseline:={AllC,AllC,AllD}. We plot a baseline to provide an objective comparison, i.e. what two cooperative strategies can achieve against AllD. Parameters: $N = 50$, $q = 0.9$, $b = 5$, $\epsilon = 0.05$. If the selection strength is very low, the baseline population can reach $Copop = 2/3$. However, the cooperation will drop to around zero for $s > 0.1$. Six of the original leading eight actually exceed that baseline, O_1, O_2 and O_7 by a large margin, but O_{3-5} also show significant improvement. Pleasing improves cooperation further in six cases and partly in another one.

of populations of interest with a baseline population BASE. Our BASE consists of the same number of strategies, but all cooperative strategies are replaced by AllC (so multiple instances of it are possible). In this BASE, no strategy uses any mechanism to support cooperation. So, we can measure the level of cooperation that results only from the composition of strategies and stochastic evolution. For example, for very low s, Copop in BASE is exactly 2/3, a substantial value. But, it will drop as the selection strength becomes larger. So, instead of choosing an arbitrary threshold s and entirely disregard any results for weaker selections, we compare the dynamics in populations with a discriminator, AllC and AllD to those in the mentioned BASE population.

Results show that the selection strength (s) has an important impact on Copop (Figure 7). Often, Copop drops and then seizes for large s. However, some pleasing strategies (such as P_6 and P_7) and an original one (O_2) show substantial Copop for the entire range (note that for some cases such as P_6, SME and hence Copop could not be computed or simulated in the way described above [2] and the values are omitted in the graph. But even the populations with O_x that eventually drop to zero show significant more cooperation than the BASE. This is especially true for O_1, O_2 and O_7. This increase is not trivial, some strategies such as P_6 show even less cooperation than the baseline for a large range of s.

But even so, pleasing is still able to make substantial further improvement in terms of the overall levels of cooperation. P_1 (standing), P_6 (stern judging) and P_7 (staying) are again the best. P_2 switches from increasing Copop to decreasing it right before $s = 1$. However, it is the only strategy, where pleasing is unsuccessful and decreases cooperation, and the failure is not repeated for other game parameters (see Figure A8). And, in four strategy populations, P_2 will always be more frequent than O_2 (Figure A9). So even if in the cases where the original have higher potential, they lose in direct competition with the pleasing strategies.

We finally study evolutionary dynamics without the restriction of rare mutations. As described above, a population may now reach any possible composition of three strategies. The frequencies of these states and the cooperation within allow us to measure Copop in a similar way as before. The full significance of the results is again better seen in comparison with the BASE population. The left column of Figure 8 shows that even such a population can have high cooperation levels if mutations are frequent and selection strength is small, whereas cooperation seizes in the opposite direction. This is to be expected but comparison with O_x populations (second column) show that they are actually slightly worse than BASE for its ideal parameter setting. But they also show broad bands of parameter ranges, where they increase Copop significantly. This is again especially true for P_1 and P_7. Both have large parameter ranges with more than 0.6 Copop, whereas rare mutations and a specific $s = 1$ had suggested, there would be less than 0.4.

The general cooperation promoting effect of pleasing remains robust. In the parameter range shown, pleasing always increases Copop in comparison to the original leading eight (even P_2). P_6 benefits the most, as before. But it is also apparent that for many strategies the absolute difference of P_x from O_X is smaller than that of O_x from BASE. Original strategies sometimes increase Copop greatly, yet pleasing

[2] In most of these cases, all invasion in more than one strategy was numerically computed as zero and therefore there is no single eigenvector. If invasion would be impossible, the evolutionary process would stop in an homogeneous state, giving that strategy a frequency of 1, but which strategy would depend on the initial state

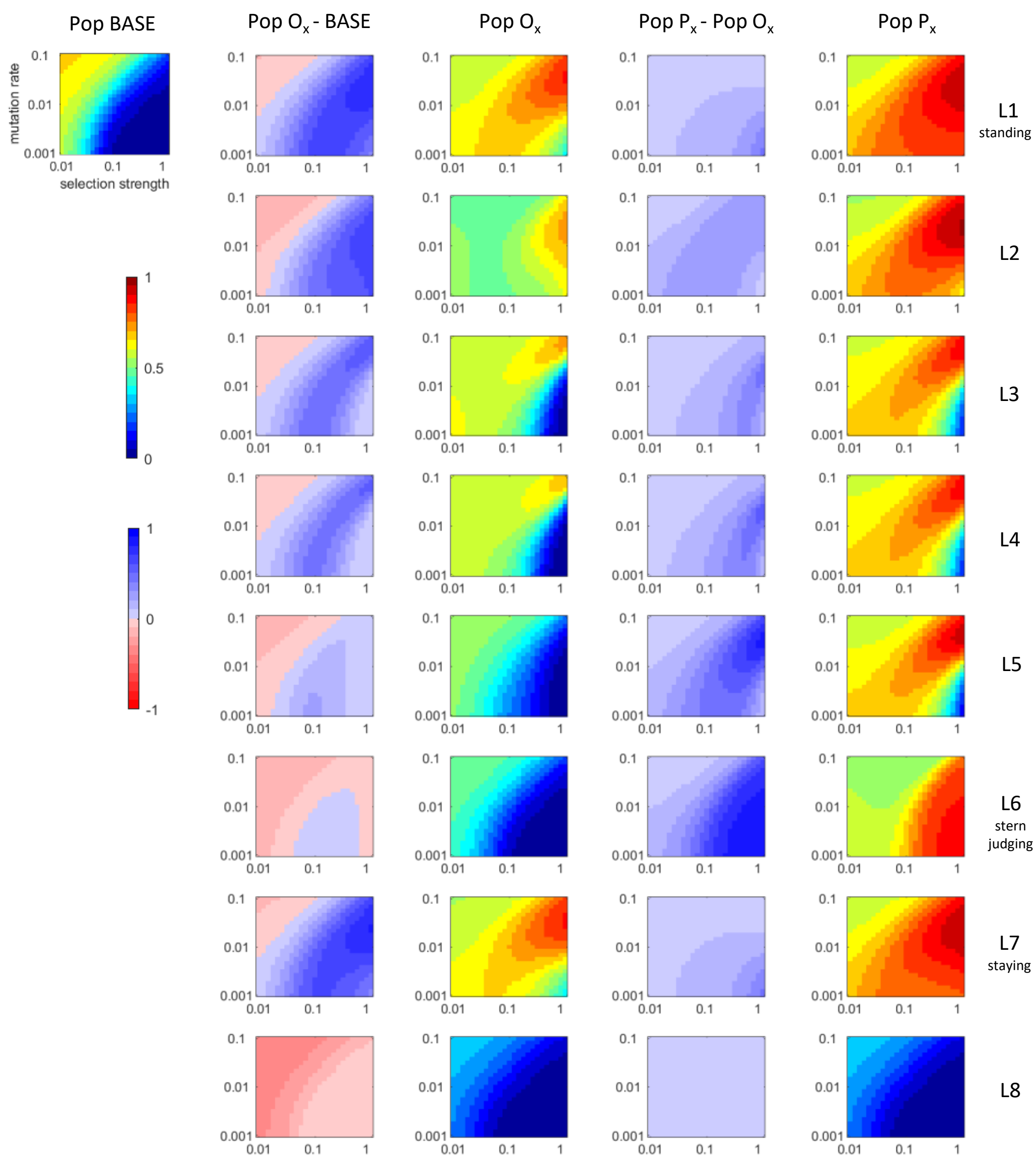

Figure 8: Space of evolutionary parameters μ (y-axis) and s (x-axis) for a wide yet not exhaustive range. Columns 1, 3 and 5 show cooperation in the population of BASE population (twice AllC and AllD), original population (O_x, AllC and AllD) and pleasing population (O_x, AllC and AllD). Columns 2 and 4 show difference of their neighboring columns (right - left) to highlight changes. Parameters: $N = 50$, $q = 0.9$, $b = 5$, $\epsilon = 0.05$.

can still lead to further improvement. A band of medium parameter sizes is best for pleasing in the same manner it is for the original. From the examined range of parameters, the scenario of rare mutation limit, with an selection strength $s = 1$ as used in previous work (Hilbe et al., 2018) (which correspond roughly to the lower right corner) provides the hardest condition for most discriminator strategies and cooperative behaviour to prevail. Their predictions of Copop might be to low and pleasing even better at enabling indirect reciprocity.

It is noteworthy that if μ is very small, populations may still spend most of the time in homogeneous states (Figure A10). The results reflect this. For $\mu = 0.001$ they resemble the predictions of the previous models for rare mutations, confirming the validity of that approach.

4 Discussion

Our analyses show three main findings. First, pleasing enhances indirect reciprocity and enables high cooperation levels, significantly outperforming the original leading eight strategies in this respect. Pleasing strategies invade original leading eight and tend to invade AllD, but can suffer from random drift towards AllC. Second, pleasing is proven efficient, working well even when pleasing only a few players or suffering an additional cost (e.g. from information acquisition). Third, considering a larger parameter space of selection strengths and mutation rates is important. Indeed, we found that, contrary to the conclusions from previous work (Hilbe et al., 2018), original leading eight, especially standing (L_1) and staying (L_7) (Sigmund, 2016), can enable high cooperation levels, exceeding those of baseline populations (Figure A2). Yet, pleasing can still lead to substantial further improvements in terms of cooperation.

4.1 Explanation of Pleasing's Success

We expected pleasing to benefit reputation-based strategies such as the leading-eight. They rely on information and pleasing uses more information for each decision, compared to the original leading-eight. However, pleasing players use this information not to update their own opinions and to directly reduce disagreement. Instead, they use it to selfishly improve their own reputations. So, it may come as a surprise that this strategy increases both stability and cooperativeness. In the study that discovered the leading-eight strategies, all successful norms had a single best set of action rules (Ohtsuki and Iwasa, 2006). They can all be described as follows: defect, unless cooperation would give you a better reputation than defection. In that sense, these strategies were acting selfish already. Selfishness can lead to stable cooperation.

Since pleasing is a better strategy to optimize one's reputation than the original leading-eight, better reputations and hence more cooperation within L-X players come not as a surprise. However, pleasing has three less direct effects. It does also raise the reputation of AllC and lowers the reputation of AllD. As a result, pleasing strategies fare much better against AllD, but lose some advantages over AllC. This causes the loss of stability for L2 for some parameter ranges. In all other cases however, it is a net win in SME.

The changes in reputation cause a further effect. Pleasing shifts reputations away from mixed (half bad, half good opinions) towards good for L-X and AllC and towards bad for AllD. As a consequence, there is less possibility for disagreement. If reputation is unanimously good (or bad), there cannot be any disagreements. Increasing reputation that was already above 50% good tends to decrease the number of disagreements between agents. Disagreement was an important cause of bad reputation for L-X in other studies with private assessment (Brandt and Sigmund, 2004; Hilbe et al., 2018). Hence, pleasing likely leads to a positive feedback to improve reputation of L-X players even further.

Third, pleasing affected the actions of agents directly and enabled more cooperation for the same amount of good reputation. Whereas an agent with 60% good reputation would receive a benefit 60% of the time from original leading eight agents, it would receive it 100% of the time from pleasing agents (if they please all other L-X in the population). A random opinion would be good 60% of the time, but the majority opinion is always good. The same principle causes the opposite effect for agents with bad reputations (i.e. AllD in most cases), who earned smaller benefits than for the same reputation with the leading eight strategies. This polarization of cooperation and defection appears to have enhanced the stability of pleasing players as well as their cooperation with each other.

As a consequence of these opinion and cooperation dynamics, pleasing achieves three things. First, it can successfully discriminate AllD, withholding help to them whilst continue to help themselves. It can do this much better than the original leading eight. Second, it maximizes its own reputation for low errors and therefore does not allow AllC to earn higher reputations or higher benefits. Third, high error rates lead to equally lower reputations of both pleasing and AllC, so pleasing will sometimes save the cost of cooperation without receiving less benefits than AllC. This way, it can achieve slightly higher payoffs. In contrast, pleasing usually fails, when invading AllD becomes very hard. Although AllD cannot invade it in turn, even neutral drift from pleasing to AllC causes a cycle that reduces the abundance of pleasing (see Figure 4 and A4).

4.2 Relevance, Limitations and Future Work

Pleasing is a novel information sharing approach to solve the problem of private assessment. It does not aim to solve the problem by transforming private into public assessment, for example by a mechanism to synchronizing opinions. Instead it offers players the ability to maximize their own reputation. Yet, it can lead to reliable indirect reciprocity even for the hardest conditions studied to far, when information is private, noisy and incomplete. We expanded the work on pleasing by studying its success against original leading eight strategies as well as the unconditional ones. With this we have substantially extended and generalised the previous findings in Krellner and Han (2020).

Several important results in the literature of indirect reciprocity were obtained using a deterministic evolutionary game approach (see a review in (Okada, 2020a)), e.g. by using replicator dynamics (Nowak and Sigmund, 1998b; Uchida, 2010; Okada, 2020b) or evolutionary stable strategy analyses (Ohtsuki, 2004; Ohtsuki and Iwasa, 2006). These works thus did not account for the impact of stochastic factors such as selection strength and mutation or behavioural exploration. Both of them have been shown to play an important role in driving the outcomes of evolutionary dynamics (Sigmund, 2016; Traulsen et al., 2010). Their values are usefully specific to concrete interaction situations and populations, which can be estimated in behavioral experiments (Traulsen et al., 2010; Rand et al., 2013; Zisis et al., 2015; Domingos et al., 2020). Indeed, when mutation is quite rare, given a similar payoff structure and behavioural update rule (i.e. Fermi), an intermediate selection strength approximately in the range $[0.01, 0.1]$ is usually observed that best explains experimental results (Rand et al., 2013; Zisis et al., 2015). We observe that, when the intensity of selection is within that range, some original leading eight can already promote the evolution of cooperation under private assessment, which is contrary to the conclusions by Hilbe et al. (2018), likely because their analysis focuses on a rather high selection strength (namely, $s = 1$) and they did not compare the cooperation levels to a baseline.

We also show that the differences in performance between original and pleasing leading eight strategies are considerable, putting our study in line with research that reports similar differences (Hilbe et al., 2018; Brandt and Sigmund, 2004) and that places certain strategies above others. Those are standing (L_1), one of the first strategies ever being investigated (Leimar and Hammerstein, 2001; Panchanathan and Boyd, 2003), stern judging (L_6) (Pacheco et al., 2006; Santos et al., 2018)) and staying (L_7) (Okada et al., 2017a; Sasaki et al., 2017)). L_1 and L_7 are somewhat similar. They are already reasonably abundant and cooperative in homogeneous states before the introduction of pleasing. With pleasing, they are abundant and cooperative over large parameter ranges. Stern judging on the other hand has no chance to evolve without pleasing. But with it, its abundance is near 100% over a considerable range of parameters and is also the most robust against additional costs of pleasing.

Is pleasing realistic? It seems highly impracticable to literally run around and ask what others think in a situation where one is supposed to help someone, especially if help is required fast. However, pleasing could be done in real life by other means. Instead of asking right beforehand, one could ask habitually and have the information ready at hand when needed. This way the information might be outdated, but a small decrease in accuracy should not diminish the effect of pleasing all together, comparable to pleasing a subset of players. Secondly, instead of asking, opinions could be inferred via theory of mind or other cognitive abilities, see e.g. (Tomasello et al., 2005; Han et al., 2012, 2011; Han, 2013). This is a natural step in studying indirect reciprocity under private assessment given that humans (and many other animals) clearly possess such cognitive abilities (Meltzoff, 2007; Woodward et al., 2009). Similar capabilities were already shown to improve indirect reciprocity, but when attributed to the observer instead of the donor (Radzvilavicius et al., 2019).

A minor limitation of the current study is that the pleasing described here will not always find the action with the best possible reputation. Rather, it is a cheap heuristic to find a very good action but could be improved. Instead of using the average opinions (see Methods) to decide for an action, it could use each opinion of each agent itself to infer the consequences of each action for each pleased agent and then choose the action that grants the best possible reputation. Considering this more sophisticated approach could push the limits of pleasing even further.

Another open question is how pleasing can deal with irrelevant or miss-leading information. Here, we did not assume that pleasing players can differentiate between their own kind and others. If those strategies would share opinions, a pleasing player would consider it valuable information. He may fail to recognize the majority opinion of pleasing players and therefore may fail to earn a good reputation. Pleasing unconditional players instead does not yield any benefit, since they treat him independently of

low error rate

		Invader			
		O_1	P_1	AllC	AllD
Resident	O_1	-	0.1119	0.0953	0.0000
	P_1	0.0000	-	0.0199	0.0000
	AllC	0.0123	0.0224	-	0.6678
	AllD	0.0442	0.0686	0.0000	-

	SME
O_1	0.0357
P_1	0.7782
AllC	0.0269
AllD	0.1592

high error rate

		Invader			
		O_1	P_1	AllC	AllD
Resident	O_1	-	0.0985	0.0160	0.0000
	P_1	0.0000	-	0.0182	0.0000
	AllC	0.1191	0.0431	-	0.6678
	AllD	0.0001	0.0047	0.0000	-

	SME
O_1	0.0060
P_1	0.2404
AllC	0.0054
AllD	0.7482

Table A2: Abundance of strategies (SME, right) and fixation probabilities of a single invading mutant in a homogeneous resident population in a population of four strategies O_1, P_1, AllC and AllD. We consider two conditions, top row: $\epsilon = 0.05$ (low error rate), bottom row: $\epsilon = 0.25$ (high error rate). In both conditions: $N = 50$, $q = 0.9$, $b = 5$, $s = 1$. Values correspond to graphical depiction in Figure 4.

his actions. However, information sharing of such players may take many forms. Since unconditional players do not actually track opinions, information they share would be some form of lying. They might always say everybody is good, or bad, or produce random answers. Other discriminators may also share information, which could be their true opinion or also lies. The effect is unclear. Some preliminary simulations show for example, that AllD players, who claim everybody is bad, actually benefit pleasing players. And, original leading eight that share their opinions with pleasing players, seem to bring them virtually no benefits. Such behavior should help out pleasing players, since they would want to please originals to gain more benefits from them. However, it seems that pleasing was already so good in invading original leading eight, that it made no difference to the results.

The full range of lying strategies will be investigated in the future. Next, we aim to study how pleasing fares against other strategies that are specifically equipped to take advantage of pleasing. Those include lying strategies as described above and also strategies that rely on the opinions of others but do not want to share their own opinions to avoid costs. The latter were pointed out as a problem of information sharing in general (Suzuki and Kimura, 2013). As mentioned, a solution could be to pay players for their information. The options seem endless.

5 Acknowledgements

TAH is supported by a Leverhulme Research Fellowship (RF-2020-603/9) and the Future of Life Institute (grant RFP2-154).

6 Appendix

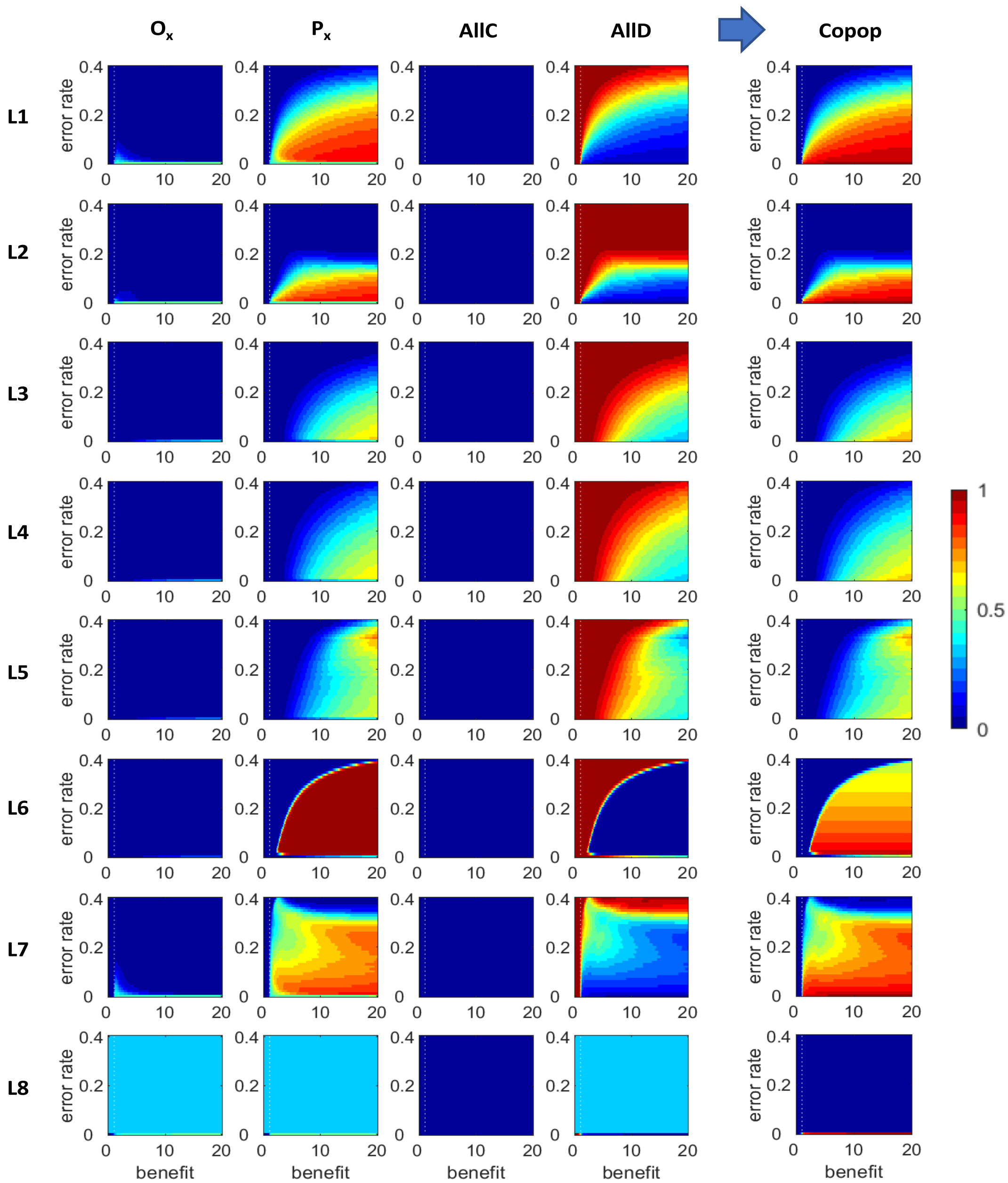

Figure A1: Abundance of strategies in populations consisting of O_x, P_x, AllC and AllD (four left columns) and resulting cooperation in the population (Copop, right column). Parameter space of benefit to cost ratio (x-axis) and rate of observation error (y-axis). Parameters: $N = 50$, $q = 0.9$, $s = 1$. For most conditions, populations consist almost entirely of P_x and AllD. Copop closely resembles the abundance of P_x, since it is the only cooperative strategy with relevant abundance and the cooperation rate in its homogeneous state is about 100% (Figure 3). The only exception is L_6, where P_6 is very abundant but less cooperative as errors become more frequent.

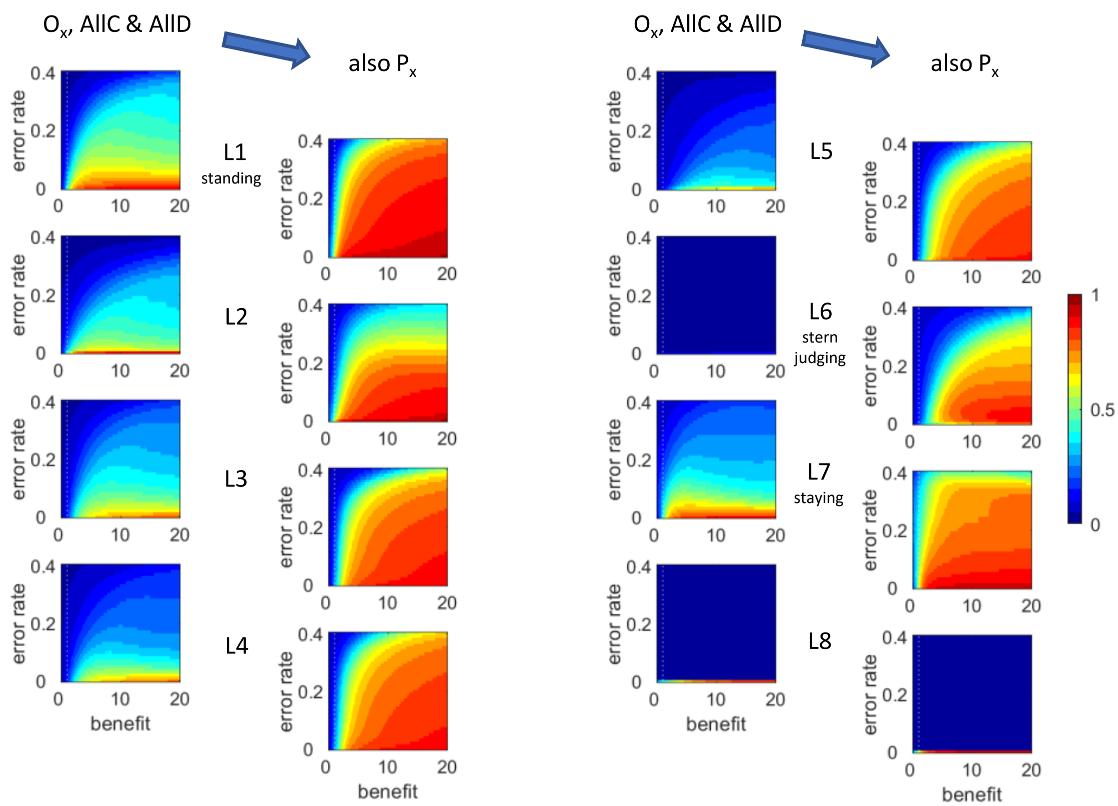

Figure A2: Results for an alternative selection strength (namely, $s = 0.1$), compared to Figure 2 where $s = 1$. Effect of pleasing on the evolution of cooperation (Copop), across benefit b (x-axis) and error rate ϵ (y-axis). First and third columns show Copop in populations of original leading-eight O_x, AllC and ALLD; second and fourth columns show populations augmented by a fourth strategy P_x. Other parameters: $N = 50$, $n_w = 50$ (i.e. pleasing all P_x players in the population), $q = 0.9$. Most original leading eight populations have substantial areas with mediocre Copop. Pleasing is able to increase Copop to above 70% for a large parameter range all but one strategy. Both populations have higher cooperation rates than when $s = 1$.

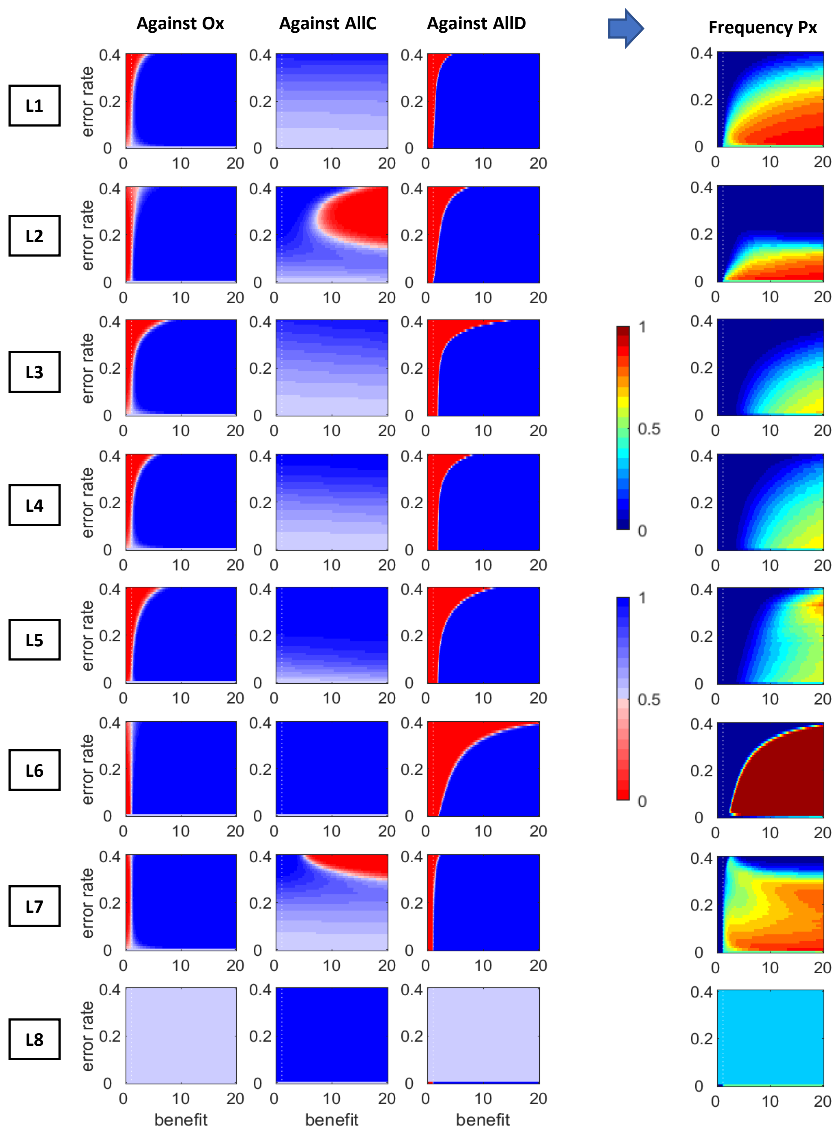

Figure A3: Relative probability of invasion of P_x against other strategies $\frac{\rho_{A,B}}{\rho_{A,B}+\rho_{B,A}}$. This probability being greater than 0.5 (shades of blue) indicates that P_x is risk dominant (three left columns). The frequency of P_x in populations with all four strategies is given in the rightmost column, both for varying the benefit to cost ratio (x-axis) and observation error rate (y-axis). Other parameters: $N = 50$, $n_w = 50$, $q = 0.9$, $s = 1$. Pleasing is risk dominant against original leading eight unless errors frequent and benefits are low (especially $b < 1$). Risk dominance can explain frequency in the case of L_6. And it predicts some low frequencies of P_x, for example when AllC is risk dominant against P_7 (high error rates, high benefits). But, risk dominance alone does not lead to a high frequency. For example, the relative probability to invade AllC increases for P_1 as error rate increases, but its frequency decreases. Absolute fixation probability is more indicative of high frequency (see Figures 4 and A4)

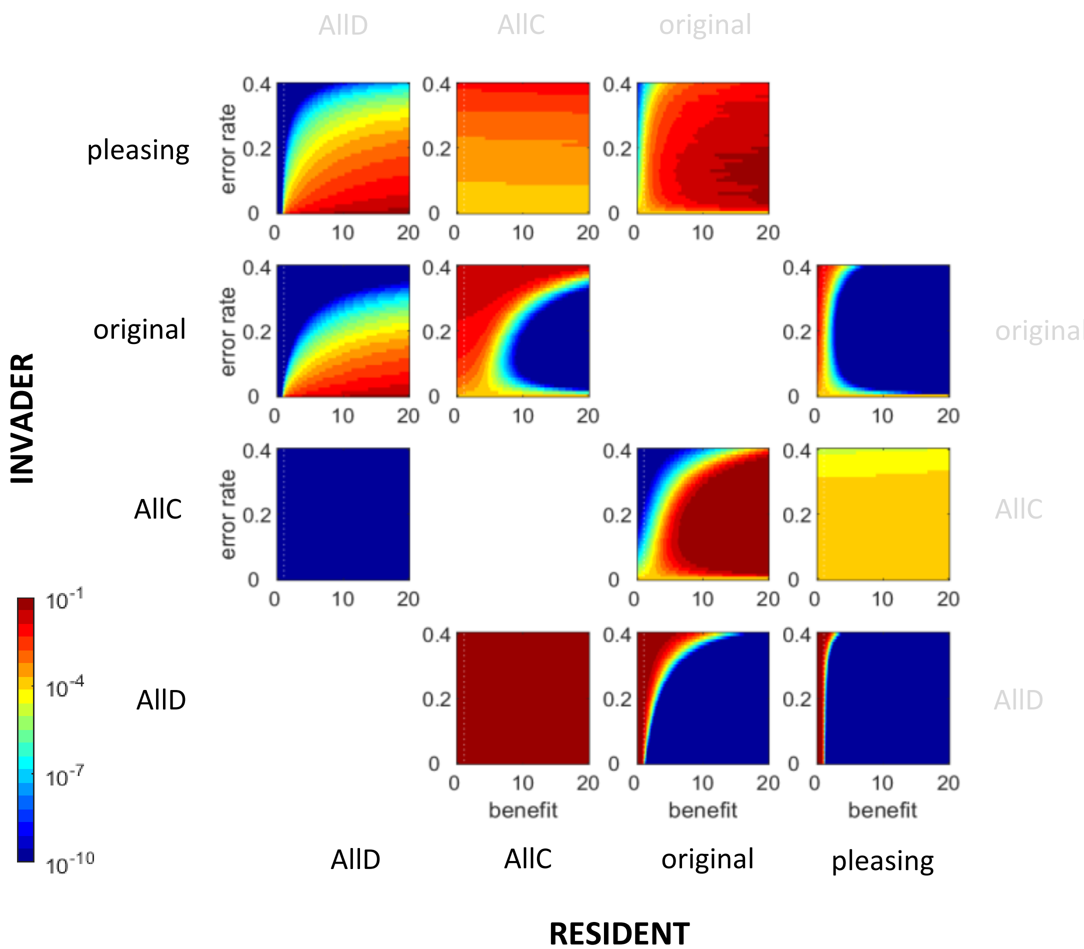

Figure A4: Fixation probabilities for leading eight number one (standing), i.e. in a population of P_1, O_1, AllC and AllD, for varying the benefit to cost ratio (x-axis) and observation error rate (y-axis). Other parameters: $N = 50$, $n_w = 50$, $q = 0.9$, $s = 1$. If fixation of P_1 in AllD becomes lower than the fixation of AllC in P_1, P_1 loses abundance to AllD, since the fixation of AllD in AllC is always very high (see Figures 4 and A1).

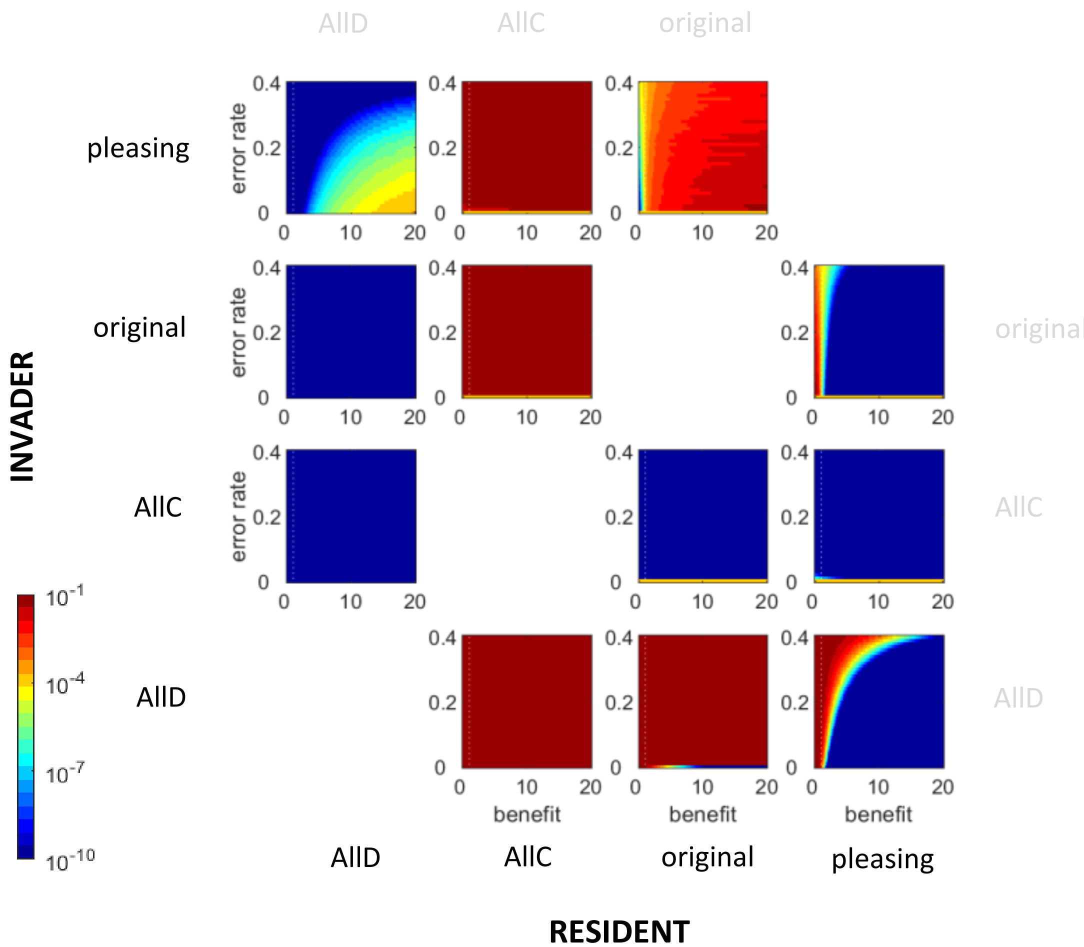

Figure A5: Fixation probabilities for leading eight number six (stern judging), i.e. in a population of P_6, O_6, AllC and AllD, for varying the benefit to cost ratio (x-axis) and observation error rate (y-axis). Other parameters: $N = 50$, $n_w = 50$, $q = 0.9$, $s = 1$. P_6 can invade AllD with a relatively small probability, i.e. at most about the chance of random drift $1/N = 10^{-3.91}$ (orange) and even smaller as benefits decrease and errors become more frequent. But in this parameter space, P_6 cannot be invaded by any other strategy and hence has an almost maximal abundance (see Figure A1).

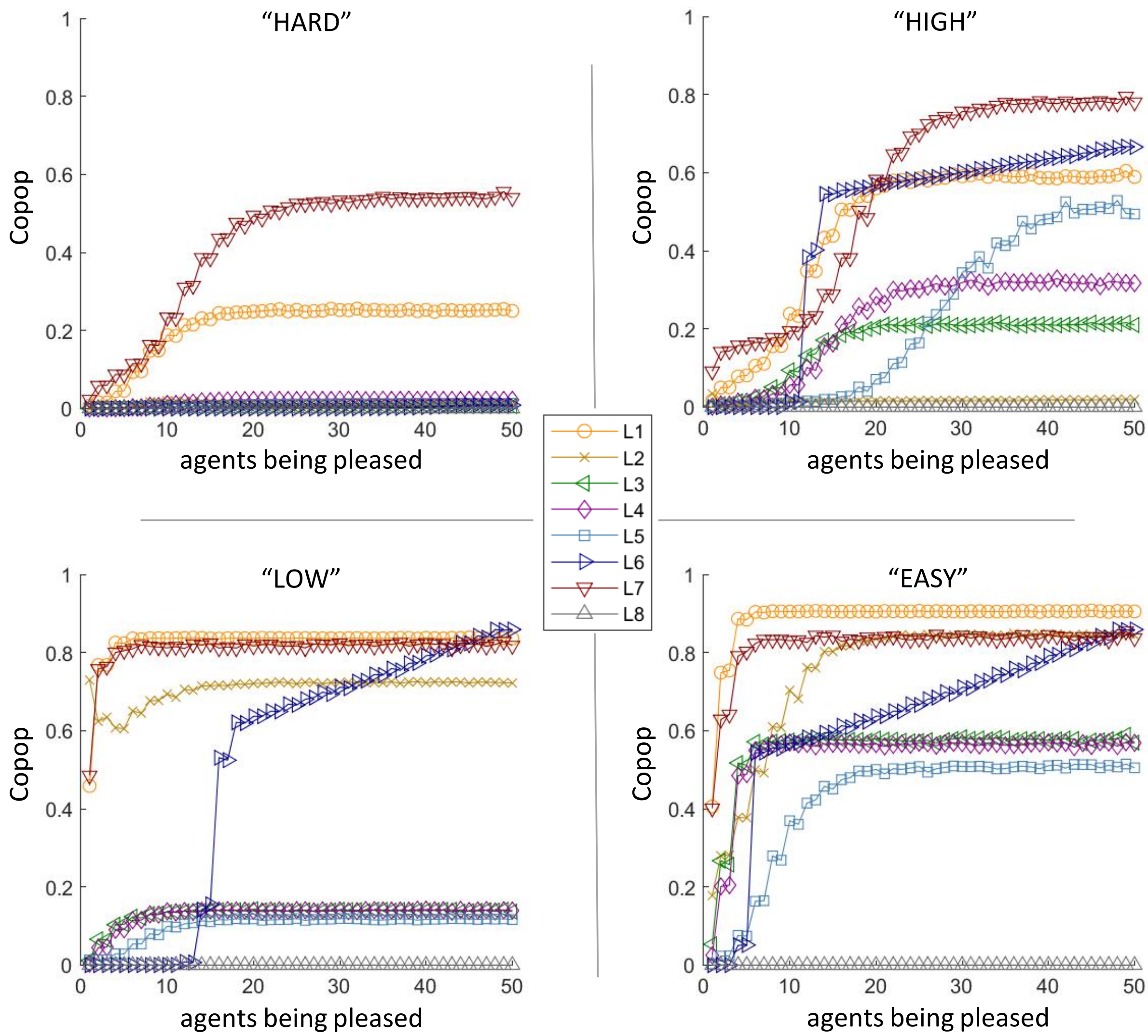

Figure A6: Evolution of cooperation in the whole population (Copop) for different effort levels of pleasing, i.e. the number of agents that the donor tries to please. We consider four conditions, low: $b = 5$, $\epsilon = 0.05$, easy: $b = 15$, $\epsilon = 0.05$, hard: $b = 5$, $\epsilon = 0.25$ and high: $b = 15$, $\epsilon = 0.25$. Other parameters: $N = 50$, $q = 0.9$, $s = 1$. As more players are considered, Copop often increases steeply and then levels of. If errors are more frequent (upper graphs), this plateauing occurs later. P_6 is an exception in that Copop steadily increases due to increasing cooperation in its homogeneous state (see Figure 3).

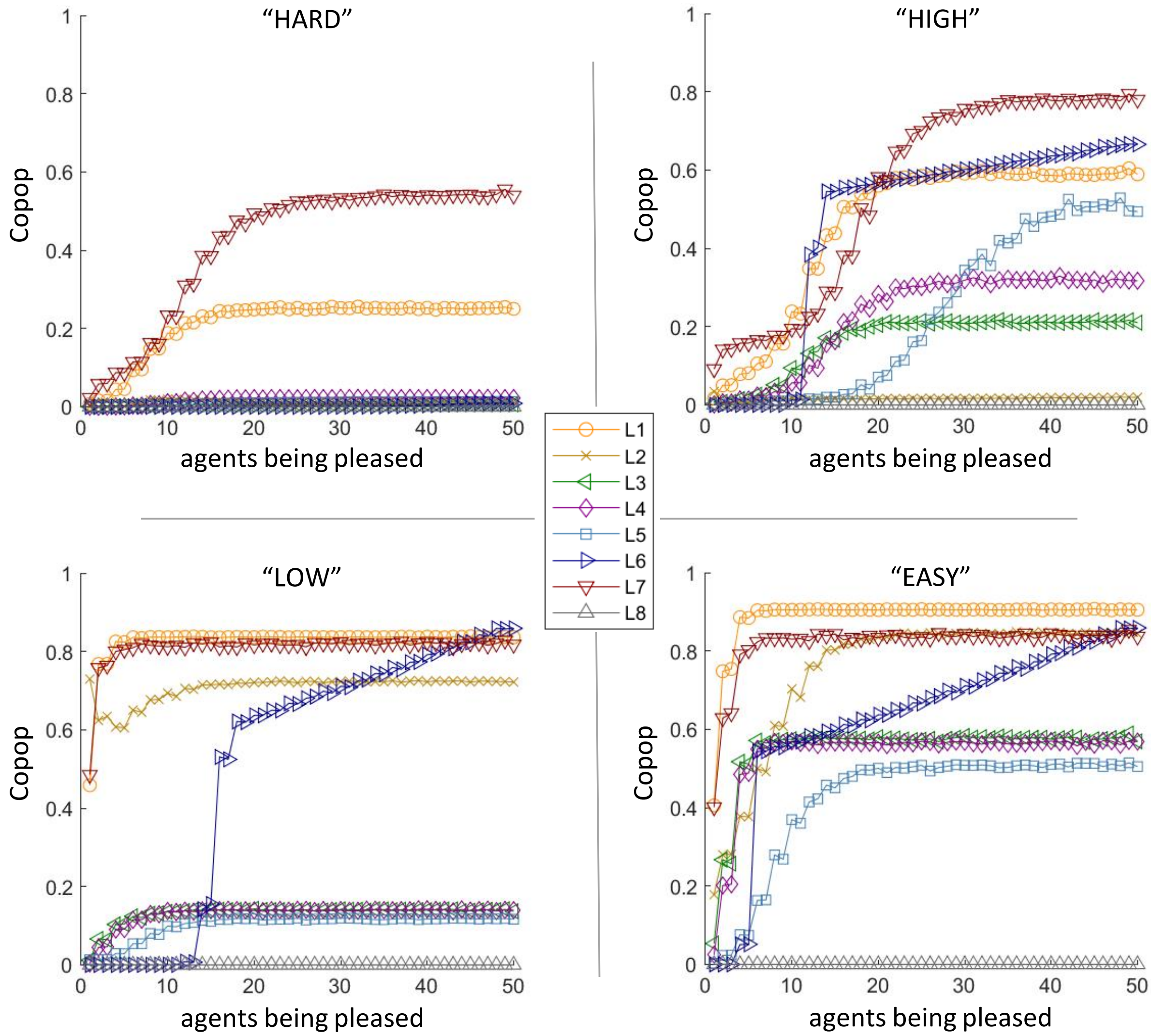

Figure A7: Effect of increasing the additional costs for pleasing on the cooperation in the whole population (Copop). Four conditions, low: $b = 5$, $\epsilon = 0.05$, easy: $b = 15$, $\epsilon = 0.05$, hard: $b = 5$, $\epsilon = 0.25$ and high: $b = 15$, $\epsilon = 0.25$. Other parameters: $N = 50$, $q = 0.9$, $s = 1$. Higher benefits (lower graphs) allow for higher Copop. For high benefits (right graphs), P_6 (stern judging) can sustain substantial costs of pleasing without a reduction in Copop.

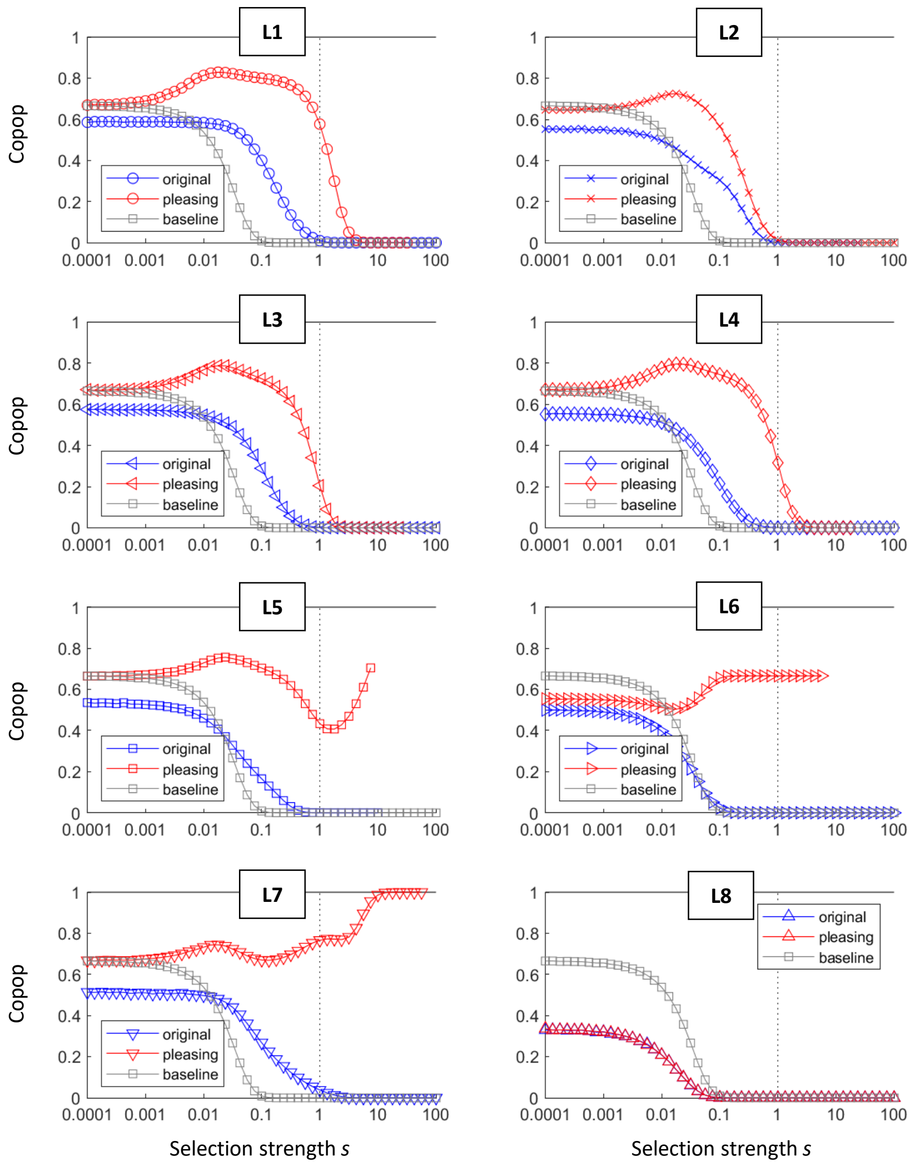

Figure A8: Alternative game parameters to those in Figure 7, namely $b = 15$, $\epsilon = 0.25$. We study the evolution of cooperation (Copop) for varying the selection strength s, in three different populations. In particular, pleasing:={AllC,AllD,P_x}, original:={AllC,AllD,O_x}; baseline:={AllC,AllC,AllD}. We plot a baseline to provide an objective comparison, i.e. what two cooperative strategies can achieve against AllD. Other parameters: $N = 50$, $q = 0.9$, $b = 5$, $\epsilon = 0.05$. If the selection strength is very low, the baseline population can reach Copop $= 2/3$. However, the cooperation will drop to around zero for $s > 0.1$. Six of the original leading eight exceed that baseline, O_1, O_2 and O_7 by a large margin, but O_{3-5} also show significant improvement. Pleasing improves cooperation further in seven cases.

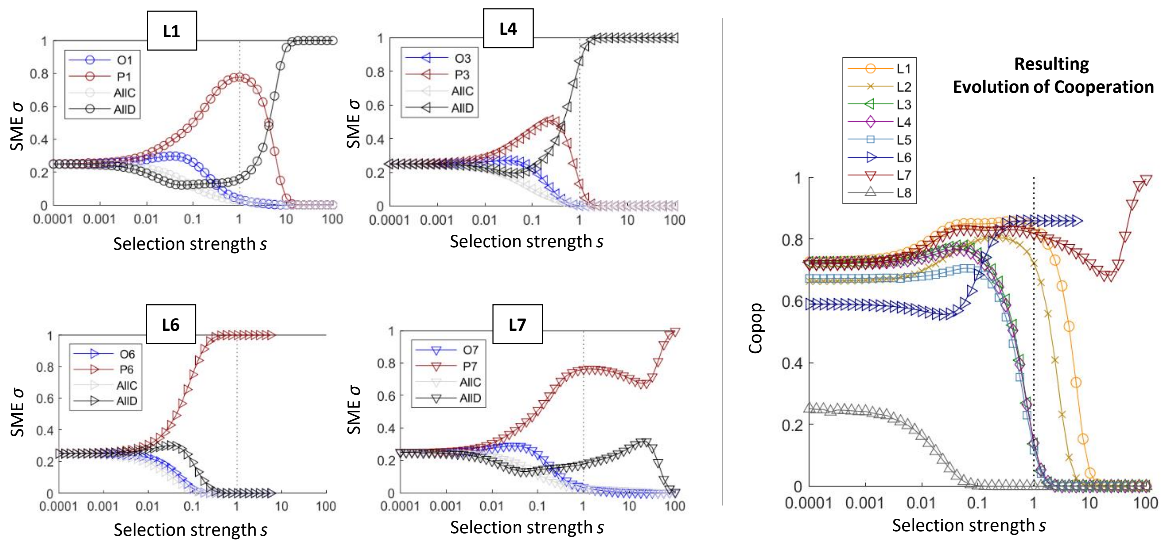

Figure A9: The effect of varying selection strength (s) on strategy frequency (left) and evolution of cooperation (right). Other parameters: $N = 50$, $q = 0.9$, $b = 5$, $\epsilon = 0.05$.

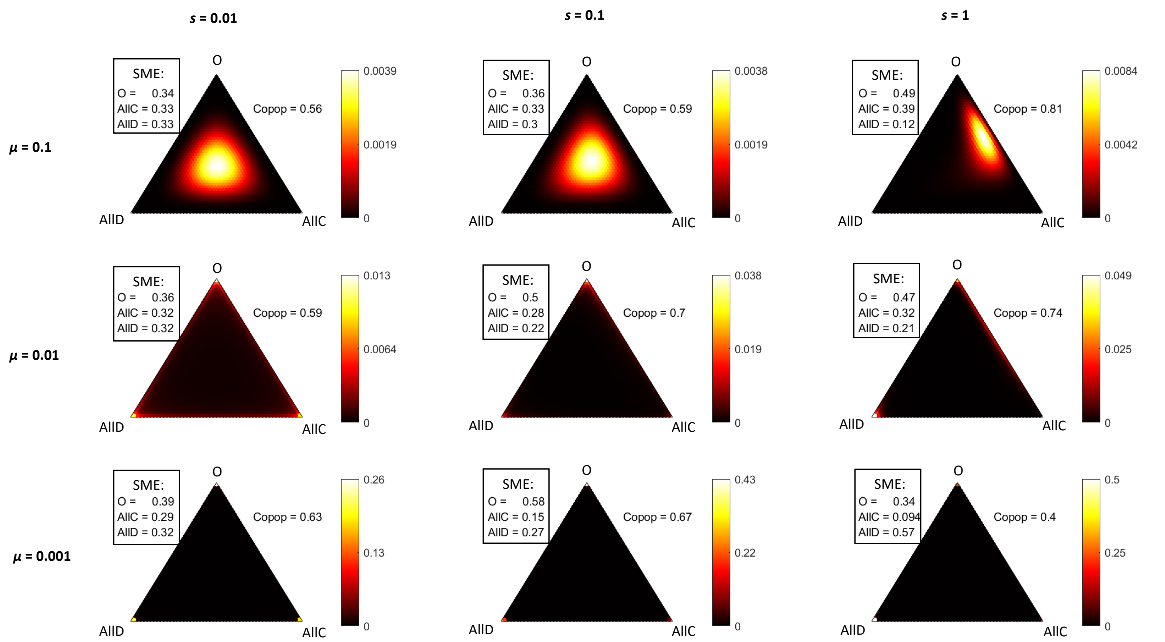

Figure A10: Heatmaps of population states, O_1 (top of triangle), AllC (right corner) and AllD (left corner). Columns show three different selection strengths, from left to right $s = 0.01,\ 0.1,\ 1$. Rows show three different mutation rates, from bottom to top $= 0.001,\ 0.01,\ 0.1$ (same range as Figure 8). Colors are customized for each graph from zero to maximum, see color bars. Parameters: $N = 50$, $q = 0.9$, $b = 5$, $\epsilon = 0.05$.
For low s, all strategies have similar abundance, as expected. For low μ, population spends most of the time in homogeneous states at the corners; and for high μ in the center.

References

Andras, P., Esterle, L., Guckert, M., Han, T. A., Lewis, P. R., Milanovic, K., Payne, T., Perret, C., Pitt, J., Powers, S. T., et al. (2018). Trusting intelligent machines: Deepening trust within socio-technical systems. *IEEE Technology and Society Magazine*, 37(4):76–83.

Antal, T., Traulsen, A., Ohtsuki, H., Tarnita, C. E., and Nowak, M. A. (2009). Mutation-selection equilibrium in games with multiple strategies. *Journal of theoretical biology*, 258(4):614–622.

Brandt, H. and Sigmund, K. (2004). The logic of reprobation: Assessment and action rules for indirect reciprocation. *Journal of Theoretical Biology*, 231(4):475–486.

Conte, R. and Paolucci, M. (2002). *Reputation in artificial societies: Social beliefs for social order*, volume 6. Springer Science & Business Media.

Domingos, E. F., Grujić, J., Burguillo, J. C., Kirchsteiger, G., Santos, F. C., and Lenaerts, T. (2020). Timing uncertainty in collective risk dilemmas encourages group reciprocation and polarization. *Iscience*, 23(12):101752.

Duong, M. H. and Han, T. A. (2020). On Equilibrium Properties of the Replicator–Mutator Equation in Deterministic and Random Games. *Dynamic Games and Applications*, 10(3):641–663.

Fehr, E. and Gachter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

Fudenberg, D. and Imhof, L. A. (2006). Imitation processes with small mutations. *Journal of Economic Theory*, 131(1):251–262.

García, J. and Traulsen, A. (2012). The structure of mutations and the evolution of cooperation. *PloS one*, 7(4):e35287.

García, J. and Traulsen, A. (2019). Evolution of coordinated punishment to enforce cooperation from an unbiased strategy space. *Journal of the Royal Society Interface*, 16(156).

Hamilton, W. (1964). The genetical evolution of social behaviour. i. *Journal of Theoretical Biology*, 7(1):1 – 16.

Han, T. A. (2013). *Intention recognition, commitment and their roles in the evolution of cooperation: From Artificial intelligence techniques to evolutionary game theory models.* Springer SAPERE.

Han, T. A. (2016). Emergence of social punishment and cooperation through prior commitments. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 2494–2500.

Han, T. A., Pereira, L. M., and Santos, F. C. (2011). The role of intention recognition in the evolution of cooperative behavior. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Two*, pages 1684–1689.

Han, T. A., Pereira, L. M., and Santos, F. C. (2012). Corpus-based intention recognition in cooperation dilemmas. *Artificial life*, 18(4):365–383.

Han, T. A., Pereira, L. M., Santos, F. C., and Lenaerts, T. (2013). Good agreements make good friends. *Scientific reports*, 3(2695).

Hilbe, C., Schmid, L., Tkadlec, J., Chatterjee, K., and Nowak, M. A. (2018). Indirect reciprocity with private, noisy, and incomplete information. *Proceedings of the National Academy of Sciences*, page 201810565.

Jiang, S.-X. and Li, J.-Z. (2007). A reputation-based trust mechanism for p 2 p e-commerce systems. *Ruan Jian Xue Bao(Journal of Software)*, 18(10):2551–2563.

Jones, A. J., Artikis, A., and Pitt, J. (2013). The design of intelligent socio-technical systems. *Artificial Intelligence Review*, 39(1):5–20.

Krellner, M. and Han, T. A. (2020). Putting oneself in everybody's shoes - Pleasing enables indirect reciprocity under private assessments. pages 402–410.

Langton, C. G. (1997). *Artificial life: An overview*. Mit Press.

Leimar, O. and Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1468):745–753.

Meltzoff, A. N. (2007). The 'like me'framework for recognizing and becoming an intentional agent. *Acta psychologica*, 124(1):26–43.

Nesse, R. M. (2001). *Evolution and the capacity for commitment*. Foundation series on trust. Russell Sage.

Nowak, M. A. (2006). Five Rules for the Evolution of Cooperation. *Science*, 314(5805):1560–1563.

Nowak, M. A., Sasaki, A., Taylor, C., and Fudenherg, D. (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428(6983):646–650.

Nowak, M. A. and Sigmund, K. (1998a). Evolution of indirect reciprocity by image scoring. *Nature*, 393(6685):573–577.

Nowak, M. A. and Sigmund, K. (1998b). The Dynamics of Indirect Reciprocity. *Journal of Theoretical Biology*, 194:561–574.

Nowak, M. A. and Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437(1291-1298).

Ohtsuki, H. (2004). Reactive strategies in indirect reciprocity. *Journal of Theoretical Biology*, 227(3):299–314.

Ohtsuki, H. and Iwasa, Y. (2004). How should we define goodness? - Reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology*, 231(1):107–120.

Ohtsuki, H. and Iwasa, Y. (2006). The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology*, 239(4):435–444.

Okada, I. (2020a). A Review of Theoretical Studies on Indirect Reciprocity. *Games*, pages 1–17.

Okada, I. (2020b). Two ways to overcome the three social dilemmas of indirect reciprocity. *Scientific Reports*, 10(1).

Okada, I., Sasaki, T., and Nakai, Y. (2017a). Tolerant indirect reciprocity can boost social welfare through solidarity with unconditional cooperators in private monitoring. *Scientific Reports*, 7(1):9737.

Okada, I., Sasaki, T., and Nakai, Y. (2017b). Tolerant indirect reciprocity can boost social welfare through solidarity with unconditional cooperators in private monitoring - SUPPLEMENTS. *Scientific Reports*, 7(1):1–14.

Okada, I., Sasaki, T., and Nakai, Y. (2018). A solution for private assessment in indirect reciprocity using solitary observation. *Journal of Theoretical Biology*, 455:7–15.

Okada, I., Yamamoto, H., and Uchida, S. (2020). Hybrid Assessment Scheme Based on the Stern-Judging Rule for Maintaining Cooperation under Indirect Reciprocity. *Games*, 11(1):13.

Pacheco, J. M., Santos, F. C., and Chalub, F. A. C. C. (2006). Stern-Judging: A Simple, Successful Norm Which Promotes Cooperation under Indirect Reciprocity. *PLoS Computational Biology*, 2(12):1634–1638.

Paiva, A., Santos, F., and Santos, F. (2018). Engineering pro-sociality with autonomous agents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Panchanathan, K. and Boyd, R. (2003). A tale of two defectors: The importance of standing for evolution of indirect reciprocity. *Journal of Theoretical Biology*, 224(1):115–126.

Perc, M., Gómez-Gardeñes, J., Szolnoki, A., Floría, L. M., and Moreno, Y. (2013). Evolutionary dynamics of group interactions on structured populations: a review. *Journal of The Royal Society Interface*, 10(80):20120997.

Perc, M., Jordan, J. J., Rand, D. G., Wang, Z., Boccaletti, S., and Szolnoki, A. (2017). Statistical physics of human cooperation. *Phys Rep*, 687:1–51.

Radzvilavicius, A. L., Stewart, A. J., and Plotkin, J. B. (2019). Evolution of empathetic moral evaluation. *eLife*, 8.

Rand, D. G. and Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, 17(8):413–425.

Rand, D. G., Tarnita, C. E., Ohtsuki, H., and Nowak, M. A. (2013). Evolution of fairness in the one-shot anonymous ultimatum game. *Proceedings of the National Academy of Sciences*, 110(7):2581–2586.

Santos, F. P., Santos, F. C., and Pacheco, J. M. (2018). Social norm complexity and past reputations in the evolution of cooperation. *Nature*, 555(7695):242–245.

Sasaki, T., Okada, I., and Nakai, Y. (2017). The evolution of conditional moral assessment in indirect reciprocity. *Scientific Reports*, 7(1):41870.

Sigmund, K. (2016). *The calculus of selflessness*. Princeton University Press.

Sigmund, K., De Silva, H., Traulsen, A., and Hauert, C. (2010). Social learning promotes institutions for governing the commons. *Nature*, 466(7308):861–863.

Sigmund, K., Hauert, C., and Nowak, M. (2001). Reward and punishment. *P Natl Acad Sci USA*, 98(19):10757–10762.

Smaldino, P. E. and Lubell, M. (2014). Institutions and cooperation in an ecology of games. *Artificial life*, 20(2):207–221.

Standifird, S. S. (2001). Reputation and e-commerce: ebay auctions and the asymmetrical impact of positive and negative ratings. *Journal of management*, 27(3):279–295.

Suzuki, S. and Kimura, H. (2013). Indirect reciprocity is sensitive to costs of information transfer. *Scientific Reports*, 3(1):1435.

Szabó, G. and Fáth, G. (2007). Evolutionary games on graphs. *Phys Rep*, 97-216(4-6).

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, 28(5):675–691.

Traulsen, A. and Nowak, M. A. (2006). Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences of the United States of America*, 103(29):10952.

Traulsen, A., Nowak, M. A., and Pacheco, J. M. (2006). Stochastic dynamics of invasion and fixation. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 74(1).

Traulsen, A., Semmann, D., Sommerfeld, R. D., Krambeck, H.-J., and Milinski, M. (2010). Human strategy updating in evolutionary games. *Proceedings of the National Academy of Sciences*, 107(7):2962–2966.

Uchida, S. (2010). Effect of private information on indirect reciprocity. *Physical Review E*, 82(3):036111.

Woodward, A. L., Sommerville, J. A., Gerson, S., Henderson, A. M., and Buresh, J. (2009). The emergence of intention attribution in infancy. *Psychology of learning and motivation*, 51:187–222.

Wu, B., Gokhale, C. S., Wang, L., and Traulsen, A. (2012). How small are small mutation rates? *Journal of Mathematical Biology*, 64(5):803–827.

Xu, J., García, J., and Handfield, T. (2019). Cooperation with bottom-up reputation dynamics. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pages 269–276. International Foundation for Autonomous Agents and Multiagent Systems.

Zisis, I., Di Guida, S., Han, T., Kirchsteiger, G., and Lenaerts, T. (2015). Generosity motivated by acceptance-evolutionary analysis of an anticipation game. *Scientific reports*, 5(1):1–11.